THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED BY H. C. CARVER

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Volume X

1939

THE ANNALS OF MATHEMATICAL STATISTICS

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I–IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE
OF MATHEMATICAL STATISTICS

Contents

	PAGE
A Contribution to the Theory of Self-Renewing Aggregates, With Special Reference to Industrial Replacement. ALFRED J. LOTKA	1
On the Mathematics of the Representative Method of Sampling. A. T. CRAIG	26
On a New Class of "Contagious" Distributions Applicable in Entomology and Bacteriology. J. NEYMAN	35
On Confidence Limits and Sufficiency with Particular Reference to Parameters of Location. B. L. WELCH	58
The Regression Systems of Two Sums Having Random Elements in Common. J. F. KENNEY	70
A Note on Confidence Intervals and Inverse Probability. ALBERT WERTHEIMER	74
Note on a Matching Problem. S. KULLBACK	77
Report of the Detroit Meeting of the Institute	81
Abstracts of Papers	88
Constitution and By-Laws of the Institute	91
Directory of the Institute	95

Vol. X, No. 1 — March, 1939

THE ANNALS OF MATHEMATICAL STATISTICS

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I–IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

A CONTRIBUTION TO THE THEORY OF SELF-RENEWING AGGREGATES, WITH SPECIAL REFERENCE TO INDUSTRIAL REPLACEMENT

By Alfred J. Lotka

1. **Introduction**. The analysis of problems of industrial replacement forms part of the more general analysis of problems presented by "self-renewing aggregates."[1] While the subject could, therefore, be treated in general and consequently rather abstract terms, for the purpose of exposition it will be advantageous to relate the discussion to concrete applications. These, in the past, have been mainly of two kinds, namely, first, applications to population analysis with related problems in genetics on the one hand and actuarial problems on the other; and second, applications to industrial replacement. As the fundamental setting of the two types of problems is very similar, leading in each case to certain integral equations, it will be advantageous to consider together both problems, or both phases of the general problem. This will incidentally give us an opportunity to observe the analogy, but also certain points of difference, between the two aspects of the problem.

Historically, the investigation of an actuarial problem came first. L. Herbelot[2] (1909) examined the number of annual accessions required to maintain a body of N policyholders constant, as members drop out by death. He assumes an initial body of N "charter" members at time $t = 0$, all of the same age, which for simplicity may be called age zero, since this merely amounts to fixing an arbitrary origin of the age scale. He further assumes the same uniform age at entry for each "new" member.

Then, if $p(t)$ is the probability at the age of entry of surviving t years, the survivors of charter members at time t will number $Np(t)$; and if $f(\tau)$ is the rate per head at which members drop out by death at time τ, being then immediately replaced by a new member of the fixed age of entry, then the survivors at time t of "new" members will evidently be given by

$$N \int_0^t f(\tau)p(t - \tau)\,d\tau$$

[1] I use here an English equivalent, as nearly as possible, to the German phrase "sich erneuernde Gesamtheiten," used by Swiss actuaries.

[2] Herbelot's original paper is disfigured with a number of misprints. It is essentially reproduced, with the errors corrected, in a paper by R. Risser (1912). The same treatment of the problem is also given by Zwinggi (1931) and by Schulthess (1935), (1937).

Hence, the condition for a constant membership N is

$$Np(t) + N\int_0^t f(\tau)p(t-\tau)\,d\tau = N \tag{1}$$

or

$$p(t) + \int_0^t f(\tau)p(t-\tau)\,d\tau = 1 \tag{2}$$

Differentiating with regard to t, and remembering that $p(0) = 1$, we have

$$p'(t) + \int_0^t f(\tau)p'(t-\tau)\,d\tau + f(t) = 0 \tag{3}$$

Equation (3) may be written

$$f(t) = -p'(t) - \int_0^t f(\tau)p'(t-\tau)\,d\tau \tag{4}$$

or, putting $(t - \tau) = a$

$$f(t) = -p'(t) - \int_0^t f(t-a)p'(a)\,da. \tag{5}$$

For the solution of the integral equation thus obtained Herbelot uses the method of successive differentiations,[3] duly pointing out its limitations, and applying it to several specific expressions for the survival function $p(a)$.

There is nothing in Herbelot's treatment to limit its application to living organisms. It is directly applicable to the problem of industrial replacement of an equipment comprising N original units installed at time $t = 0$, and maintained constant by the replacement of disused units with new.

Next in chronological order, of publications dealing with the type of problem with which we are here concerned, is a paper by Sharpe and Lotka (1911), who use Hertz's form of solution for the integral equation involved.[4] To this I wish

[3] This method is also followed in dealing with the problem of renewal by Risser (1912), (1920); Zwinggi (1931); Schulthess (1935), (1937); Preinreich (1938). All these authors applied their reflections to arbitrarily assumed frequency distributions for the renewal function, of simple analytical form. For example, among the more recent applications is one by Schulthess, who uses the function $p(t) = \left(1 - \frac{t}{\omega}\right)^m$; and quite recently, Preinreich has suggested the use of a Type I Pearson frequency curve on the basis of Kurtz's observational data. It is to be noted, however, that when it comes to actual application, Preinreich does not use an ordinary Pearson Type I curve nor actual observational data of any kind, but very conveniently simplifies the Pearson formula by giving integral values, namely 1 and 2, to the exponents, thereby reducing to triviality the task of applying the method of differentiation. None of these authors makes any attempt to deal with actual numerical observations which, in practice, fall far wide of any of the simple analytical formulae employed by them.

[4] P. Hertz, *Mathematische Annalen*, 1908, vol. 65, pp. 84 to 86.

to refer in some detail, adding to the original exposition in the light of later developments. The treatment of the subject proceeds here along somewhat broader lines, but, with obvious changes in the meaning of the symbols, and with certain modifications and limitations which are themselves of interest, the development is immediately applicable to economic systems composed of units having a characteristic "mortality" in use.

A population of living organisms, unlike industrial equipment, has practically no beginning. We know its existence only as a continuing process. Accordingly the equation for its development is most naturally framed without *explicit* reference to any "charter members."

The basis of the analysis is as follows:

In a population growing solely by excess of births over deaths (i.e. in the absence of immigration and emigration), the annual female births $B(t)$ at time t are the daughters of mothers a years old, born at time $(t - a)$ when the annual female births were $B(t - a)$. If fertility and mortality are constant and such that a fraction $p(a)$ of all births survive to age a, and are then reproducing at an average rate $m(a)$ daughters per head per annum, then, evidently,[5]

$$B(t) = \int_0^\infty B(t - a)p(a)m(a)\,da \tag{6}$$

$$= \int_0^\infty B(t - a)\varphi(a)\,da. \tag{7}$$

This is the fundamental equation in its original form, and, as noted above, it does not explicitly refer to any initial state, though, as will be seen presently, in order to make the problem determinate, data regarding the system at some particular period must be given. For the present we note that (7) can be written

$$B(t) = \int_t^\infty B(t - a)\varphi(a)\,da + \int_0^t B(t - a)\varphi(a)\,da \tag{8}$$

$$B(t) = B_1(t) + \int_0^t B(t - a)\varphi(a)\,da. \tag{9}$$

It is to be noted that the right hand member of (8), splits the total births $B(t)$ into two sections, those in which $(t - a) < 0$, that is, births of daughters whose mothers were born *before* $t = 0$; and those for which $(t - a) > 0$, that is births of daughters whose mothers were both *after* $t = 0$. The former section is denoted by $B_1(t)$ in (9). The function $B_1(t)$ thus defined will be found, in the

[5] Here and elsewhere in these developments the limits of the integral have, for simplicity, been written 0 and ∞. This ensures the inclusion of all nonvanishing terms in the integrand; the inclusion of terms for which either $\varphi(a)$ or $B(t - a)$ vanishes does not, of course, affect the value of the integral. If $\varphi(a)$ is represented between the limits α, ω of the reproductive period by some analytical expression, such as a Pearson frequency function, it is, of course, understood that outside the range α, ω we must put $\varphi(a) = 0$.

further development, to play a significant rôle. Here it will suffice to point out that it vanishes for all values of t greater than ω, the upper limit of the reproductive period, because $\varphi(a)$ vanishes for these values of a.

2. **Special case.** A case of special interest is that in which $B_1(t)$ represents the births of daughters whose mothers were all born in an interval of time $t = -dt$ to $t = 0$. In that case the first integral in (8) reduces to a single term, so that

$$B(t) = B(0)\varphi(t)\,dt + \int_0^t B(t-a)\varphi(a)\,da \tag{10}$$

or, putting

$$B(0)\,dt = N_0 \tag{11}$$

$$B(t) = N_0\varphi(t) + \int_0^t B(t-a)\varphi(a)\,da. \tag{12}$$

This last equation holds also if a finite number of births take place (or are regarded as taking place) at a point of time $t = 0$.

Equations (10) and (12) are of interest as basic for the examination of the progeny of an infinitesimal population element,[6] that is, of a "zero" generation, born at time zero. In that case $B_1(t)$ is the annual rate of births in the "first" generation, and is simply proportional to $\varphi(t)$, i.e.

$$B_1(t) = N_0\varphi(t) \tag{13}$$

For the sake of greater generality the development has so far been given in terms of the phenomenon of replacement (reproduction) as it presents itself in a population of living organisms. But it should be noted here that, with appropriate changes in the meaning of $\varphi(a)$, equation (12) is directly applicable to the problem of industrial renewal in an installation originally installed at some point of time and maintained at a constant level by the replacement of each unit by a new one, the moment it is disused. In that case the "rate per head of reproduction" $m(a)$ at age a is evidently the same thing as the "death rate per head" at age a, namely

$$\mu(a) = -\frac{dp(a)}{p(a)\,da} = -\frac{p'(a)}{p(a)} \tag{14}$$

so that

$$\varphi(a) = p(a)\mu(a) \tag{15}$$

becomes

$$\varphi(a) = -p'(a). \tag{16}$$

[6] A. J. Lotka, (1928), (1929).

Reverting now to the fundamental equation in its first form (6), a trial substitution

$$B(t) = Qe^{rt} \tag{17}$$

is found to satisfy this equation, provided that r is a root of the characteristic equation

$$\int_0^\infty e^{-ra} \varphi(a) \, da = 1 \tag{18}$$

We may speak of (17) as a particular solution of (6) or (7). It is easily seen that the sum of such particular solutions is also a solution, i.e.

$$B(t) = Q_1 e^{r_1 t} + Q_2 e^{r_2 t} + \cdots \tag{19}$$

where r_1, r_2 etc., are roots of the characteristic equation (18).[7]

For real values of r the function

$$F(r) = \int_0^\infty e^{-ra} \varphi(a) \, da \tag{20}$$

decreases monotonically as r increases, since, from its nature, $\varphi(a) > 0$ for all values of a. Hence (18) can have only one real root r_1, and we shall have

$$r_1 \gtreqless 0 \quad \text{according as} \quad \int_0^\infty \varphi(a) \, da \gtreqless 1. \tag{21}$$

If $u + iv$ is a complex root of (18) then

$$1 = \int_0^\infty e^{-ua} \cos va \, \varphi(a) \, da \tag{22}$$

$$0 = \int_0^\infty e^{-ua} \sin va \, \varphi(a) \, da \tag{23}$$

and it is evident from (22) that $u < r_1$, since $\cos (va) \leqq 1$ for all values of a. The real part of any complex root of (18) is, therefore, algebraically less than the real root r_1.

This reasoning[8] is evidently quite independent of the particular form of $\varphi(a)$, and is thus equally true, whether $\varphi(a)$ be given in purely empirical form (defined by a table of values), or as a standard form of frequency curve, such as for example a Pearson curve of suitable type.

The roots of (18) can be determined directly, though rather laboriously, from

[7] For a discussion of the convergence of the series (19) see G. Herglotz, *Mathem. Annalen*, 1908, vol. 65, pp. 87 et seq.

[8] Adapted from P. Hertz, *Math. Annalen*, 1908, vol. 65, pp. 1–86; G. Herglotz, ibid. pp. 87–106. The Hertz solution is also applied to a similar problem by J. B. S. Haldane, *Proc. Cambridge Phil. Soc.*, 1926, vol. 23, p. 607. A particularly detailed development is given by H. T. J. Norton, *Proc. London Math. Soc.*, 1926, vol. 28, p. 21.

equations (22) and (23); or, they can be brought into relation with the Thiele semivariants μ of the function $\varphi(a)$ defined by

$$F(r) = \int_0^\infty e^{-ra}\varphi(a)\,da = m_0 e^{-\mu_1 r + \frac{1}{2!}\mu_2 r^2 - \cdots} \tag{24}$$

where m_n is the nth moment of $\varphi(a)$ and the seminvariants μ can be computed from the moments by the algorithm

$$\begin{cases} m_1 = \mu_1 m_0 \\ m_2 = \mu_1 m_1 + \mu_2 m_0 \\ m_3 = \mu_1 m_2 + 2\mu_2 m_1 + \mu_3 m_0 \\ m_4 = \mu_1 m_3 + 3\mu_2 m_2 + 3\mu_3 m_1 + \mu_4 m_0 \end{cases} \tag{25}$$

etc.

In terms of these seminvariants the characteristic equation (18) becomes

$$\mu_1 r - \mu_2 \frac{r^2}{2!} + \cdots - \log_e m_0 = \log_e 1 = 2\pi n i \tag{26}$$

where n takes on all positive and negative integral values. Separating the real and imaginary parts in (26), and retaining seminvariants up to the fourth,

$$\begin{aligned} \psi(u, v) = \frac{\mu_4}{4!}(u^4 - 6u^2v^2 + v^4) - \frac{\mu_3}{3!}u(u^2 - 3v^2) \\ + \frac{\mu_2}{2!}(u^2 - v^2) - \mu_1 u + \log_e m_0 = 0 \end{aligned} \tag{27}$$

$$\chi(u, v) = \frac{\mu_4}{3!}uv(u^2 - v^2) + \frac{\mu_3}{3!}v(v^2 - 3u^2) + \mu_2 uv - \mu_1 v = 2\pi n. \tag{28}$$

If $\varphi(a)$ does not differ too widely from the normal (Gaussian) distribution, so that seminvariants of higher than second order can be neglected for roots in the neighborhood of $u = 0$, $v = 0$, we shall have, approximately[9]

$$\frac{\mu_2}{2!}(u^2 - v^2) - \mu_1 u + \log_e m_0 = 0 \tag{29}$$

$$\left(u - \frac{\mu_1}{\mu_2}\right)v = \frac{2\pi n}{\mu_2} \tag{30}$$

[9] The relations which follow hold *exactly* if $\varphi(a)$ is actually a normal curve. It should be noted, however, that this can not be strictly the case, since the infinite tail of the curve on the negative side would imply replacement or reproduction antedating the original installation or zero generation. Nevertheless, a normal frequency curve will be admissible if the part of the curve extending into the negative age field is negligible. For a concrete example (electric light bulbs) see E. J. Gumbel, "Die Verteilung der Gestorbenen um das Normalalter," *Aktuarske Vedy* (Praze), 1933, p. 90.

or, putting

$$\left(u - \frac{\mu_1}{\mu_2}\right) = U \tag{31}$$

we have

$$U^2 - v^2 = \left(\frac{\mu_1}{\mu_2}\right)^2 - \frac{2 \log_e m_0}{\mu_2} \tag{32}$$

$$Uv = \frac{2\pi n}{\mu_2}. \tag{33}$$

It is thus seen that in these circumstances the roots u, v correspond to the points of intersection of the hyperbola (32) centered at $u = \frac{\mu_1}{\mu_2}$, $v = 0$, with a family of hyperbolas (33) concentric with (32), but with their axes at 45° to those of (32).

The intersections of the hyperbolas (33) with the axis of v are given by putting $u = 0$ in (30), namely

$$v = \frac{2\pi n}{\mu_1} \tag{30a}$$

This also gives, approximately, the frequency of the oscillatory components for which u is sufficiently small. In particular, for the first component, we have, in that case

$$v = \frac{2\pi}{\mu_1} \tag{30b}$$

so that its wave length is (approximately) μ_1, the mean of the $\varphi(a)$ curve.

These facts are illustrated in Fig. 1, drawn to scale according to the vital statistics of the United States, 1920, for which the requisite computations were available from prior publications (Lotka, (1928), (1929)). The diagram is drawn in full, showing four intersections of each hyperbola of the family (33). Actually values of v occur in pairs, corresponding to conjugate roots $u \pm iv$. The intersections in the two upper quadrants must be disregarded, as they do not correspond to roots of (18).

To simplify notation let us write (32), (33) in the form

$$U^2 - v^2 = K \tag{32a}$$

$$Uv = C. \tag{33a}$$

Solving for U^2, v^2 we find

$$U^2 = \tfrac{1}{2}\{K \pm \sqrt{K^2 + 4C^2}\} \tag{34}$$

$$v^2 = \tfrac{1}{2}\{-K \pm \sqrt{K^2 + 4C^2}\} \tag{35}$$

from which, incidentally, it is seen that

$$U^2 + v^2 = \sqrt{K^2 + 4C^2} \tag{36}$$

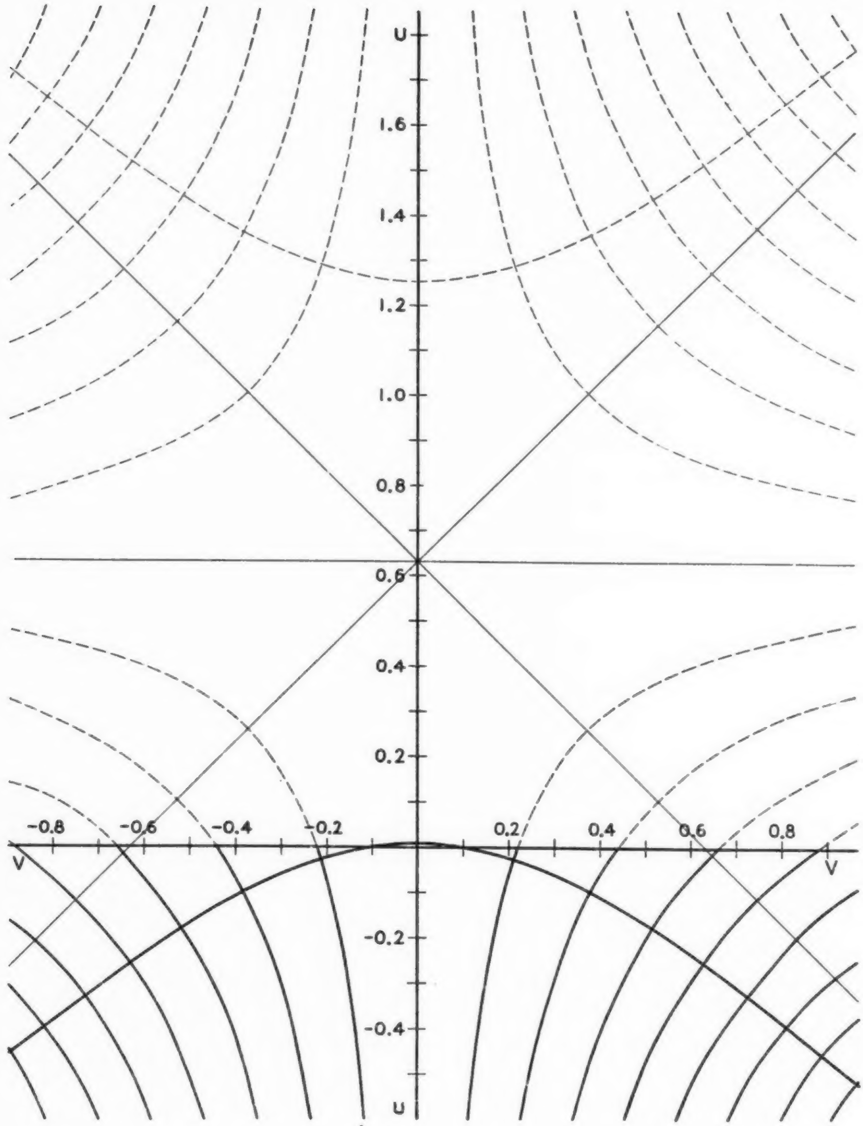


FIG. 1. ROOTS OF FUNDAMENTAL EQUATION (18) AS INTERSECTIONS OF CURVE (32) WITH FAMILY OF CURVES (33)

and hence, that the intersections of the hyperbola (32) with (33) lie on circles of radius

$$R = \sqrt[4]{K^2 + 4C^2}. \tag{37}$$

When the third and fourth moments (and therefore third and fourth seminvariants) are taken into account[10] the hyperbolas become distorted into new curves, though the general topographic features of the diagram tend to be preserved. In particular, the property of orthogonality of intersection of the curves (32) with (33) is preserved, in accordance with a well-known property of conjugate functions.[11] This is shown in the left hand panel of Fig. 2, drawn for the same data as Fig. 1, but including not only the hyperbolic curves, but also the corresponding modified curves obtained by retaining the third and fourth seminvariants in the computation.[12] Only the quadrant relevant to the location of the roots is shown.

3. **The coefficients Q in the solution** (19). These are determined by initial conditions, being, in fact related to the function $B_1(t)$. As their determination in the original paper by Hertz and Herglotz is rather complicated, the following relatively simple method, resembling that by which the constants in a Fourier series are determined, is of interest:

Multiplying equation (9) by $e^{-r_i t}$, where r_i is a root of (18), transposing terms, and integrating between the limits 0 and ω, where ω is the highest age for which $\varphi(a)$ has a value other than zero, we have

$$\int_0^\omega e^{-r_i t} B_1(t)\, dt = \int_0^\omega e^{-r_i t}\left\{B(t) - \int_0^t B(t-a)\varphi(a)\, da\right\} dt. \tag{38}$$

Introducing the solution (19) in the right hand member of (38), we obtain

$$\int^\omega e^{-r_i t} B_1(t)\, dt = \sum Q_j \int_0^\omega e^{-r_i t}\left\{e^{r_j t} - \int_0^t e^{r_j(t-a)}\varphi(a)\, da\right\} dt \tag{39}$$

$$= \sum P_{ij} \qquad (j = 1, 2, 3, \cdots). \tag{40}$$

Consider now a particular term P_{ij} in the sum $\sum$. Multiplying out the exponentials we obtain

$$P_{ij} = Q_j \int_0^\omega e^{-(r_i - r_j)t}\left\{1 - \int_0^t e^{-r_j a}\varphi(a)\, da\right\} dt \tag{41}$$

which, in view of the characteristic equation (18) reduces to

$$P_{ij} = Q_j \int_0^\omega e^{-(r_i - r_j)t} \int_t^\omega e^{-r_j a}\varphi(a)\, da\, dt \tag{42}$$

$$= Q_j \int_0^\omega e^{-r_j a}\varphi(a) \int_0^a e^{-(r_i - r_j)t}\, dt\, da. \tag{43}$$

Hence, if $i \neq j$

$$P_{ij} = \frac{Q_j}{r_j - r_i}\int_0^\omega e^{-r_j a}\varphi(a)\{e^{-(r_i - r_j)a} - 1\}\, da \tag{44}$$

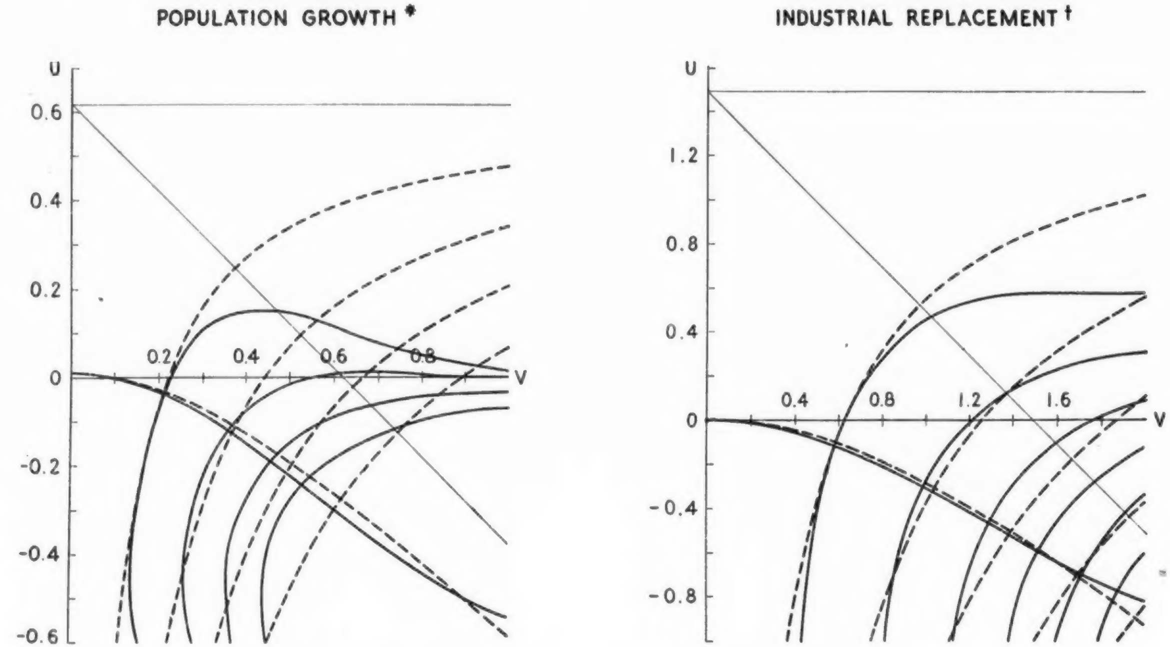
---

[10] Which is as far as curve fitting by Pearson's method goes.

[11] See, for example, W. E. Byerly, *Integral Calculus*, 1888, p. 289.

[12] For a given value of u equation (27) is a biquadratic in v, and equation (28) is a cubic in v lacking the second degree term. The computation of the curves is in consequence relatively simple.

ROOTS OF FUNDAMENTAL EQUATION (18) AS INTERSECTIONS OF CURVE (27) OR (32) WITH FAMILY OF CURVES (28) OR (33) RESPECTIVELY



—— Computed on basis of *first four* seminvariants, fourth degree equations (27), (28)
- - - - Computed on basis of *first two* seminvariants, equation of hyperbolas (32), (33)
† Data from Kurtz, E.B., "Life Expectancy of Physical Property", 1930, page 104, fig. 50.
* Data from Lotka, A.J., "The Progeny of a Population Element," American Journal of Hygiene, 1928, page 875

FIG. 2

$$= \frac{Q_j}{r_j - r_i} \left\{ \int_0^\omega e^{-r_i a} \varphi(a)\, da - \int_0^\omega e^{-r_j a} \varphi(a)\, da \right\} \tag{45}$$

$$= 0 \tag{46}$$

since r_i and r_j are both roots of (18). But if $i = j$, then (44) is of the indeterminate form 0/0 and we must refer back to equation (43), from which, with $i = j$, we obtain, instead of (44) a different expression, namely

$$P_{ii} = Q_i \int_0^\omega e^{-r_i a} \varphi(a) \int_0^a dt\, da \tag{47}$$

$$= Q_i \int_0^\omega a e^{-r_i a} \varphi(a)\, da \tag{48}$$

so that the only term in the sum $\sum$ in equation (40) that does not vanish is the term P_{ii} and finally

$$Q_i = \frac{P_{ii}}{\int_0^\omega a e^{-r_i a} \varphi(a)\, da} \tag{49}$$

$$= \frac{\int_0^\omega e^{-r_i t} B_1(t)\, dt}{\int_0^\omega a e^{-r_i a} \varphi(a)\, da} \tag{50}$$

$$= \frac{\int_0^\omega e^{-r_i t} \left\{ B(t) - \int_0^t B(t-a) \varphi(a)\, da \right\} dt}{\int_0^\omega a e^{-r_i a} \varphi(a)\, da} \tag{51}$$

or, finally, in view of (20)

$$Q_i = \frac{\int_0^\omega e^{-r_i t} \left\{ B(t) - \int_0^t B(t-a) \varphi(a)\, da \right\} dt}{-\{F'(r)\}_{r=r_i}}. \tag{52}$$

The coefficients Q are thus fully determined by (50) or its equivalents (51) or (52), when initial conditions are given, that is, when the function $B_1(t)$ is given for $0 < t < \omega$ or, what amounts to the same thing, when $B(t)$ is known for this range of values of t. For complex roots the denominator in (52) becomes,[13] in view of (27), (28)

$$-\frac{dF(r)}{dr} = -\left\{ \frac{\partial \psi}{\partial u} + i \frac{\partial \chi}{\partial u} \right\} = G - iH \tag{53}$$

---

[13] Since r_i is a root of $F(r) = 1$, we have

$$\left[\frac{dF(r)}{dr}\right]_{r=r_i} = \left[\frac{dF(r)}{F(r)\, dr}\right]_{r=r_i} = \left[\frac{d \log_e F(r)}{dr}\right]_{r=r_i}$$

where G and H can be expressed in terms of the seminvariants by partial differentiation of (27), (28) with regard to u, namely

$$G = \mu_1 - \mu_2 u + \frac{\mu_3}{2!}(u^2 - v^2) - \frac{\mu_4}{3!}(u^3 - 3uv^2) + \cdots \tag{54}$$

$$H = \mu_2 v - \mu_3 uv + \frac{\mu_4}{3!}(3u^2 v - v^3) - \cdots . \tag{55}$$

In the special case that the "zero generation" is composed of N_0 individuals (or "units") all born (or "entering") at time zero, the coefficients Q are correspondingly simplified in form. For the term in the real root r we have

$$Q = \frac{N_0}{-F'(r)} . \tag{56}$$

Conjugate complex root terms unite in pairs,[14] giving

$$Q'e^{(u+iv)t} + Q''e^{(u-iv)t} = \frac{2N_0 e^{ut}}{G^2 + H^2}\{G \cos vt - H \sin vt\}. \tag{57}$$

Unless $\varphi(a)$ is a normal distribution, the computation of the roots, u, v, and the coefficients G, H, in terms of seminvariants becomes impracticable for higher order roots, which then have to be computed directly and laborously from equations (22), (23). In practice components of very high order will hardly be needed, nor will their use be warranted, since the high order seminvariants, which are then involved, are not usually known with sufficient accuracy. An exception occurs when the $\varphi(a)$ curve is essentially of the nature of a composite curve. This is what actually happened in the case of the curve of reproduction for a human population. For details on this point the reader must be referred to my paper "The Progeny of a Population Element".

4. **Alternative Representation of the Function** $B(t)$. By the application of the Hertz-Herglotz solution of the integral equation (6), the evolution of a population or aggregate is represented as the resultant of a series of damped oscillations.

Additional insight into the nature of the renewal process is gained by viewing the total renewals as composed of contributions from successive "generations".[15]

[14] For details see A. J. Lotka, *The Progeny of a Population Element*, p. 892.

[15] In the case of a population the term "generation" calls for no explanation: mother, daughter and granddaughter, for example, represent three generations; in the case of industrial replacement, the term is to be understood in this sense, that the original installation constitutes the original or zero generation, the units introduced to replace disused units of the zero generation constitute the "first" generation, renewal of these the second, and so on.

This explanation may seem unnecessary. However, from some correspondence received by the writer it seems that perhaps some readers have confused the generations thus defined with successive "cycles" of duration equal to the extreme "length of life" of the units. With such "cycles" we are not here concerned.

This leads to an alternative representation, in which the evolution of the aggregate appears as the sum of a series of frequency curves, each corresponding to the contribution of one generation to the total births or replacements at time t.[16]

In order to realize this second representation we note, first of all, equation (7) applies not only to the total births at time t, but, with slight modification, also to the births in any particular generation. Here it will be convenient to consider the special case of a zero generation of N_0 individuals (or units) all born (or installed) at time $t = 0$.

The births (or renewals) in the "first" generation, that is offspring of the zero generation, or renewals of disused units of the zero generation, will be distributed in time according to the equation

$$B_1(t) = N_0\varphi(t). \tag{58}$$

For the second generation, or renewals of disused units of the first generation, we shall have

$$B_2(t) = \int_0^t B_1(t - a)\varphi(a)\, da \tag{59}$$

[16] This alternative approach of the problem bears some superficial resemblance to a method followed by R. Frisch in his article "Sammenhengen mellem primaerinvesteringen og reinvestering" (*Statsekonomisk Tidskrift*, 1927, p. 117). Frisch also follows up the distribution in time of first, second, and higher order replacements, and gives diagrams bearing a superficial resemblance to Fig. 4 in the present text. But Frisch's development has otherwise little in common with that here presented. He deals with equipment composed of various units, with expectation of life varying discontinuously or continuously from one unit to another, but fixed at a single value for a given unit. To use one of his own examples, it is as if a wooden hammer with a life of one year were always replaced by another wooden hammer, also with a life of one year, and so on: while a steel hammer, with a life of three years, were always replaced by another steel hammer, also with a life of exactly three years. The analogous case in population analysis would be presented by a population in which length of life were strictly hereditary, so that a man dying at age 50 would have a son, grandson, etc., each dying at age 50. In the field of industrial replacement and in population analysis alike this is a highly unrealistic supposition.

Needless to say, with these basic assumptions, Frisch's resulting equations differ fundamentally from those here given, and the distribution curves for successive orders of replacements, as shown in Frisch's Fig. 3 do *not* have the property that the j-th seminvariant of the k-th order replacement curve is k times that of the j-th seminvariant of the first order curve, except for $j=1$. The fact is that Frisch's curves in his Fig. 3 are all similar, except for a constant factor applied to the vertical scale and its reciprocal applied to the horizontal scale. In this case all the corresponding seminvariants, except the first, are evidently unchanged in passing from one curve to the next. Frisch, as a matter of fact, does not introduce seminvariants into his discussion at all. The Hertz solution he could not possibly introduce, since his fundamental equations are not of a form appropriate for the use of the Hertz solution.

The later sections of Frisch's paper deal with somewhat more complicated cases, but they all involve the assumption of "strict heredity," that is, the assumption that a unit with length of life v is replaced by another having exactly the same length of life v. At any rate, that is the understanding I have formed of the Danish text, studied with the assistance of a native of Scandinavia. All the formulae in the text bear out this understanding.

and, generally, for the $(j + 1)$th generation[17]

$$B_{j+1}(t) = \int_0^t B_j(t - a)\varphi(a)\, da. \tag{60}$$

Now, by a well-known property[18] of the Thiele seminvariants, it follows from (58), (59), (60), that the seminvariants of the distribution-in-time of the births (or replacements) in the jth generation are simply the j-tuple of the corresponding seminvariants of the first generation, that is, of $\varphi(t)$.

Furthermore, it is easily shown that as j, the order of generation, increases, the distribution of renewals approaches[19] the normal (Gaussian) frequency distribution.

By virtue of these properties the distribution curves for successive generations are easily constructed.[20]

The sum total of the contributions of successive generations should, of course, agree with the expression for the total annual births $B(t)$ at time t given by the fundamental equation (9). In point of fact, by summing the left and the right hand members of equations (58), (59), and (60) for all generations up to the highest, say the n-th, "reproducing" at time t, we find

$$B(t) = \sum_{j=1}^{j=n+1} B_j(t) = B_1(t) + \int_0^t \sum_{j=1}^{j=n} B_j(t - a)\varphi(a)\, da. \tag{61}$$

Since the n-th is the highest generation contributing,[21] the value of the integral in (61) is not changed by writing n instead of $n + 1$ as the upper limit of the summation sign on the right. But then (61) becomes simply

$$B(t) = B_1(t) + \int_0^t B(t - a)\varphi(a)\, da$$

---

[17] The births in the j-th generation extend at most from $t = j\alpha$ to $t = j\omega$, but it is not necessary to take this into account in writing the limits of the integrals in (60) and corresponding equations, because the inclusion or exclusion of vanishing terms in the integrand does not affect the value of the integral. Similar remarks apply to the effect of the limited range of $\varphi(a)$. See also footnote 5.

[18] For details, see A. J. Lotka, "The Progeny of a Population Element," *American Journal of Hygiene*, 1928, vol. 8, p. 875; also "The Spread of Generations" *Human Biology*, 1929, vol. 1, p. 305.

[19] In practice quite rapidly, even if $\varphi(a)$ is far from normal.

[20] For the case in which $\varphi(a)$ is a Pearson Type I curve, details of the process are given in my paper "Industrial Replacement," *Skandinavisk Aktuarietidskrift*, 1933, p. 51. I may here remark that such a Pearson Type I curve for the distribution in the first generation does not strictly give again a Pearson Type I curve in the second generation, because the moments beyond the 4th are neglected in fitting such a curve. But it must be remembered that the same neglect is practiced in the original fit of the data, so that the fit in the second generation will in general be as adequate as that in the first, provided, of course, that proper attention is paid to Pearson's criteria.

[21] The special case that the limiting n so defined is ∞ would require special discussion, which, however, presents no great difficulty. As this case is of little if any practical importance, this discussion is here omitted.

that is, summation of the contributions of individual generations to the total annual births, leads us back to the fundamental equation (9), which confirms the correctness of our analysis.

TABLE I

Age Schedule of Survivorship and of Replacements[22] *in First Generation*

Age Interval	Survivors from Original Installation to Beginning of Specified Age Interval	Replacements Within Specified Age Interval
0–1	100,000	—
1–2	100,000	—
2–3	100,000	300
3–4	99,700	900
4–5	98,800	1,800
5–6	97,000	3,000
6–7	94,000	5,700
7–8	88,300	10,300
8–9	78,000	14,100
9–10	63,900	13,900
10–11	50,000	13,800
11–12	36,200	13,200
12–13	23,000	10,400
13–14	12,600	6,300
14–15	6,300	3,700
15–16	2,600	2,200
16–17	400	400
17–18	—	—

5. **Application to Kurtz's data.** An extensive collection of numerical data (mortality curves) on renewal of industrial equipment has been published by E. B. Kurtz (1930), (1931). By way of example the analysis developed above has been applied to the data "Group III," as fitted by him with a Pearson Type I curve, namely[23]

$$B_1(t) = 14,950\left(1 + \frac{t - 10}{12.67}\right)^{9.16}\left(1 - \frac{t - 10}{10.43}\right)^{7.54}. \tag{62}$$

[22] Data from E. B. Kurtz, *Life Expectancy of Physical Property*, 1930, Table 22, Cols. 5 and 6, p. 86, and p. 104, Fig. 50.

[23] The numerical values of the constants in the formula as here given differ slightly from those given by Kurtz, perhaps owing to the retention by him of higher decimals in his computations. There is also an inconsistency between Kurtz's use of 10 for the mean in his formula, whereas on his drawing the mean is placed at 100.

The aperiodic component is the number of units originally installed (arbitrarily assumed as 100,000) divided by the mean of the frequency curve (equation 62). Following Kurtz, this has also been arbitrarily made equal to 10, which simply implies a particular choice of time unit. The fundamental data and characteristics are set forth in Tables I and II. The first six oscillatory components, were computed retaining moments and seminvariants up to μ_4, with the results shown in Table III and in Figs. 2 (right hand panel), 3 and 4.

TABLE II

Moments and Seminvariants of Curve of Replacements in First Generation[24]

j	Moments[25] m_j	Seminvariants μ_j
0	100,000	
1	0	10 [26]
2	671,924	6.7192
3	130,070	−1.3007
4	12,323,200	−12.1228

TABLE III

Constants of the Series Solution (19) of Integral Equation (7) for First Six Oscillatory Components Computed from First Four Moments and Seminvariants of an Industrial Replacement Curve[27]

Order of Component n	u	v	G	H	$\frac{G}{G^2+H^2}$	$\frac{H}{G^2+H^2}$
0	0	0	10.0000	0	.10000	0
1	−.11009	.57767	11.1688	4.1458	.07869	.02921
2	−.30144	.98920	14.3353	7.6696	.05423	.02902
3	−.46500	1.28383	18.4982	10.4425	.04100	.02314
4	−.59500	1.51475	23.1094	12.7773	.03314	.01832
5	−.69800	1.70500	29.2088	14.8877	.02718	.01385
6	−.78000	1.86117	32.5165	16.7797	.02429	.01253

In particular, Fig. 4 shows the curve obtained by the summation of the first six oscillatory components superposed over the aperiodic (constant) component. It also shows the distribution curves of the first five generations within the range of the time scale on the diagram. Summation of these reproduces,

[24] Data from E. B. Kurtz, *Life Expectancy of Physical Property*, 1930, Table 22, p. 86, and Fig. 50, p. 104.

[25] Moments taken about age 10.

[26] This value of μ_1 is taken with reference to the origin.

[27] Data from E. B. Kurtz, *Life Expectancy of Physical Property*, 1930, p. 104, fig. 50.

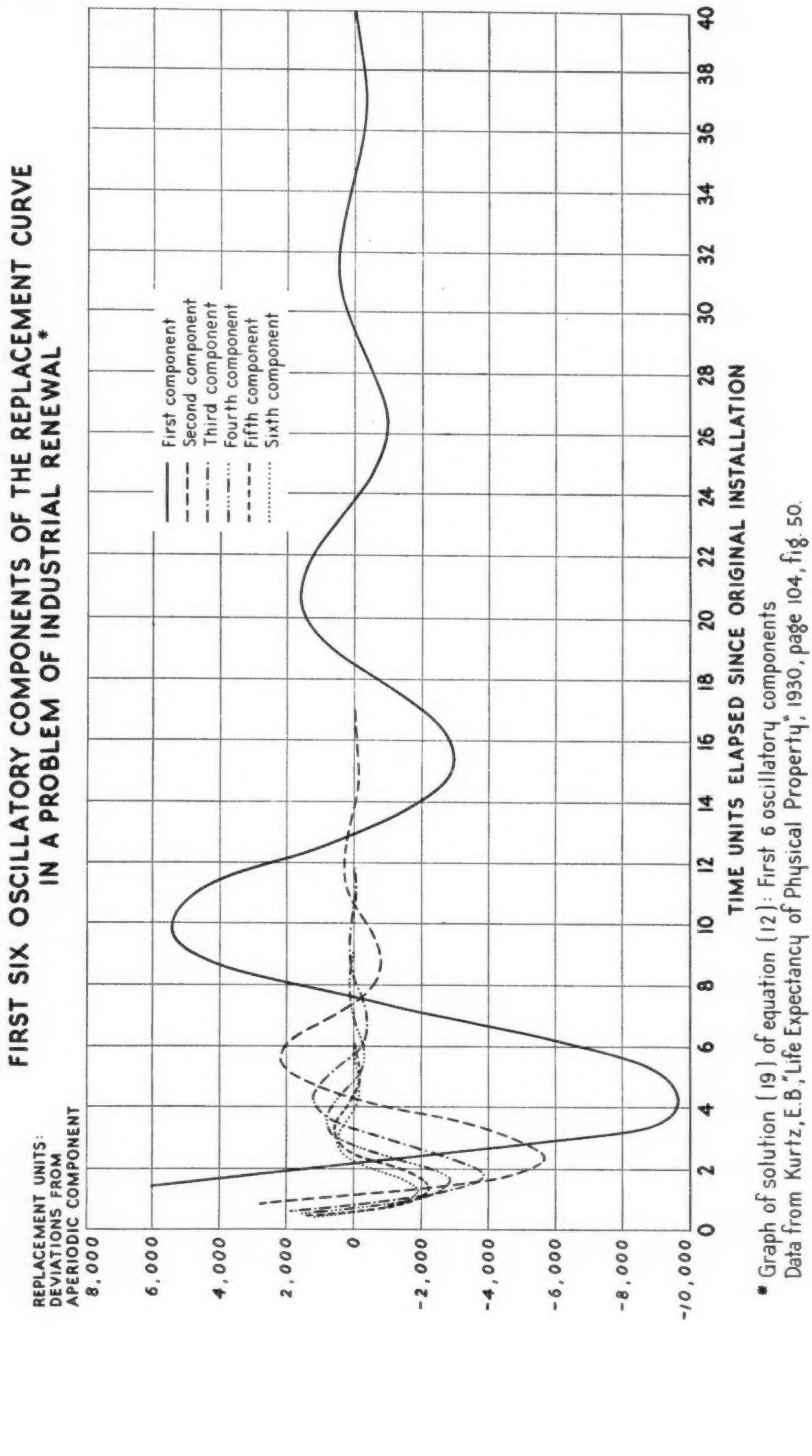
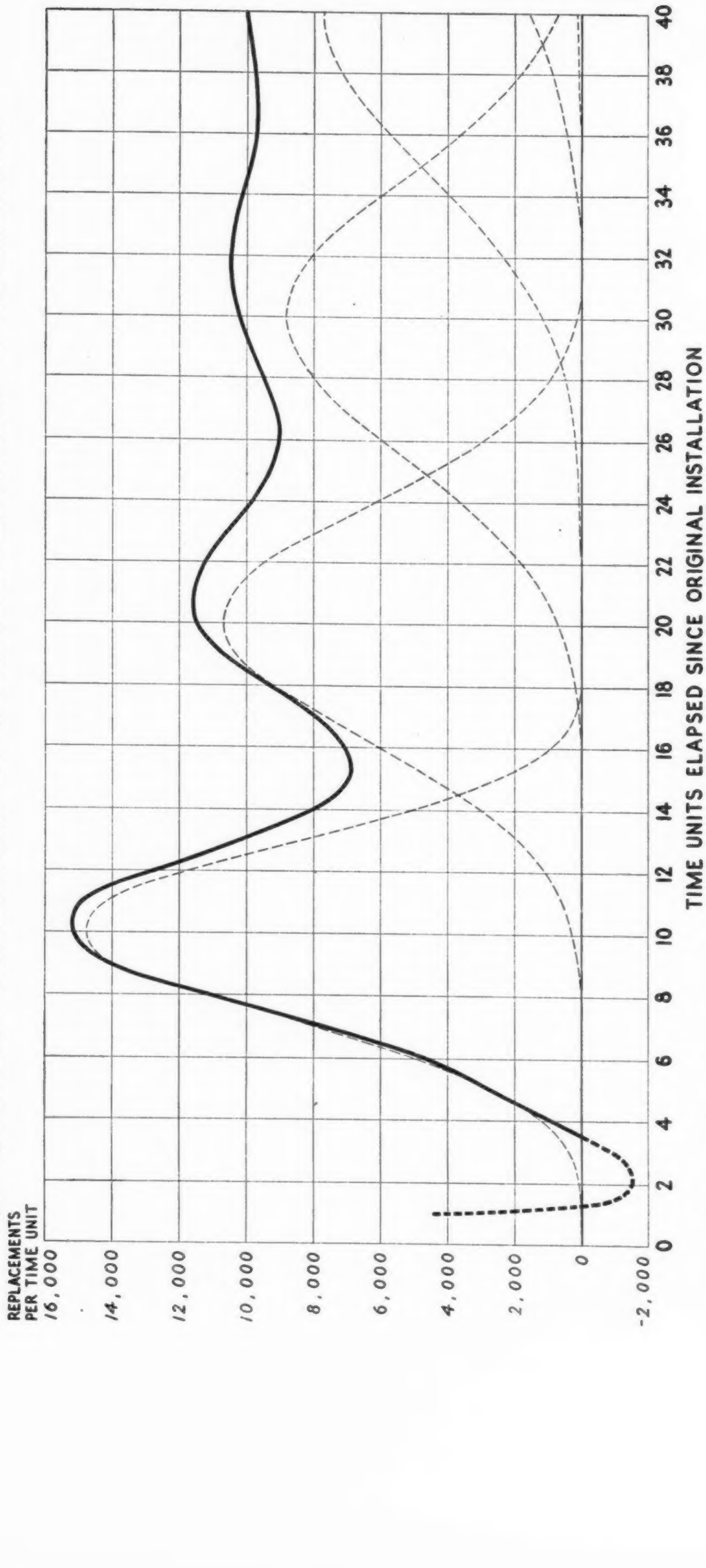

FIRST SIX OSCILLATORY COMPONENTS OF THE REPLACEMENT CURVE
IN A PROBLEM OF INDUSTRIAL RENEWAL *
REPLACEMENT UNITS: DEVIATIONS FROM APERIODIC COMPONENT
8,000
6,000
4,000
2,000
0
-2,000
-4,000
-6,000
-8,000
-10,000
First component
Second component
Third component
Fourth component
Fifth component
Sixth component
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
TIME UNITS ELAPSED SINCE ORIGINAL INSTALLATION
* Graph of solution (19) of equation (12): First 6 oscillatory components
Data from Kurtz, E.B., "Life Expectancy of Physical Property," 1930, page 104, fig. 50.

FIG. 3

SUMMATION OF APERIODIC AND FIRST SIX OSCILLATORY COMPONENTS, AND FREQUENCY DISTRIBUTIONS OF SUCCESSIVE GENERATIONS OF REPLACEMENTS *



* Graph of Solution (19) of equation (12): summation of first six oscillatory components
Data from Kurtz, E. B., "Life Expectancy of Physical Property", 1930, page 104, fig. 50.

FIG. 4

within the errors of drawing, the resultant curve of the oscillatory solution, except for the very early stages of the process, where the oscillatory solution is of no practical interest, because the first generation alone dominates the whole process, and this is given by the observational data direct or after fitting with the curve such as (62).

It remains to consider briefly the relative advantages of the method of solution by differentiation, as originally applied by Herbelot, Risser, and others, on the one hand, and the use of the Hertz-Herglotz expansion, as introduced for the treatment of this type of problem by Sharpe and Lotka.

One obvious advantage of the method of differentiation *when it is applicable*, is that the result is obtained in the form of a closed, finite expression *for each cycle*.

Against this is to be reckoned, first, that the range of application of the method is severely limited. Preinreich in a recent issue of *Econometrica* (1938) uses for an illustration of the method a Pearson Type I curve, but in the very special and trivial form that the exponents are integers, namely 1 and 2. In practice the exponents will always be fractional, and then successive differentiations do not terminate as obligingly as in Preinreich's case. As already noted, Preinreich, though citing Kurtz's observational data on industrial replacement, discreetly abstains from using these for his numerical example.

Secondly, the disadvantage of a solution in form of an infinite series is more apparent than real. In practice the first few terms of the series obtained by the Hertz-Herglotz method will usually give an adequate representation of the facts, except for a short period immediately following the first installation. It is true that here this method, unless carried to high order components, may give an imperfect representation of industrial replacements, and may, in fact, give impossible negative values in this region, as in the example exhibited in Fig. 4. But this is practically unimportant, because in practice there will actually be few, if any, such very early replacements in an installation of finite dimensions. In fact, second and higher order replacements immediately after first installation are obviously out of the question in practice. For example, it may well happen once in a while that a telegraph pole is demolished on the very first day of service by collision with a truck. It is even imaginable that its replacement, put up the same day, might again be immediately demolished. But even in a country-wide installation one would hardly expect a third, fourth or fifth replacement to be required on the day of installation. In other words, that part of the replacement curve which relates to the very early period after first installation, is composed practically of first replacements only.

So for example in the diagram, Fig. 4, the curve of total replacements, up to about $t = 8$, is simply the curve of first replacements, which is given directly by the data of the problem. Within the range of errors of drawing the influence of higher components are quite unobservable in this region.

The case is even more favorable in the application of the method to the problem of population growth, for here there is actually no reproduction what-

ever until age α (say about 15) is reached. The part of the curve defined by the series (19) carried only to a finite number of terms,[28] and applied to values of $t < \alpha$, is therefore simply rejected.[29] It may save many words of explanation if the reader is simply referred to Fig. 4 on p. 897 of my previous publication "The Progeny of a Population Element," which illustrates the point, the minimum age of reproduction being just short of 15.

A major disadvantage of the method by differentiation is that it demands that the frequency distribution function $\varphi(a)$ be given in the form of a suitable analytic expression, or if it is not so given, that a *suitable* function or curve be fitted to it. The Hertz-Herglotz method, on the contrary, is directly applicable to the *raw data, regardless of their form.* Incidentally, curve fitting as practiced by Kurtz may produce a singular result. In 6 out of 7 of his types, the fitted frequency curve extends into negative field, implying that there are some replacements even before the actual installation. This may not be a very serious defect if the area of the curve in the negative field is negligible, but it should not pass unnoticed.

One of the principal merits of the Hertz-Herglotz expansion is that it renders the course of events over their whole extent, and, in particular, makes clear the mode of approach to the ultimate state represented by the aperiodic term. Because the method by differentiation requires a separate expression for each cycle, it is at best ill adapted to present to the eye or to the mind a comprehensive view of the evolution of the aggregate as a whole.

In the introductory paragraphs it was pointed out that the problems of population growth and those of industrial replacement were closely analogous, though there were certain points of difference. It is of interest here to give consideration to these differences.

One of these has already been noted. Replacement of industrial equipment may begin from the very moment of first installation, since accident as well as wear and tear must be provided for. Organic reproduction, on the other hand, does not occur immediately after birth. One result of this is that for any finite value of t, the number of generations contributing to the total births is itself finite; on the contrary, in the case of industrial replacement, if we interpret the equation (7) literally, there are at any moment an infinite number of generations contributing. In practice this, of course, does not occur, and the equation

---

[28] There are, of course, limitations to the application of the solution (19). No one with any experience in the treatment of practical problems by mathematical analysis would think of fitting, by means of a reasonably limited number of terms, the first phases of the processes here discussed, in the case of a rectangular distribution of the first generation, for example. But the distributions with which we are actually concerned in practice are far from rectangular. Such as they are, they are well adapted to the method, as is seen in the two examples illustrated.

[29] There is nothing unusual in this rejection of negative values of the frequency function where it falls outside the range of actual values. It is what we all do in using such a frequency curve as Pearson's type I, defined by a function which becomes negative outside the range of actual interest.

does not truly represent the facts in that a continuous distribution is assumed throughout, whereas for the higher order replacements ultimately the early frequencies are so thinned out that the discreteness of the units can no longer be disregarded.

Nevertheless, from the very start we must be prepared to consider several generations of replacement as contributing to the total; this lends a certain special interest, in dealing with the first cycle of replacements, to the method of solution by differentiation, as used by Herbelot, Risser, Zwinggi, Schulthess, and lately Preinreich. It is true that this interest is much diminished by the limitations in the applicability of the method.

On the other hand, in the case of organic reproduction, for the early part of the first cycle, the progeny of a population element belongs exclusively to a single ("first") generation. Between $t = 15$ and $t = 30$, in our example, only first generation births are taking place, and here the solution (19) is of more theoretical than practical interest, since the distribution of births is simply that of the first generation births.

Another point of difference is that the curve of $\varphi(a)$ in the case of industrial replacement, if we may judge by Kurtz's data, is a comparatively well behaved Pearson type curve. On the contrary, the corresponding curve of organic reproduction is a very inconvenient type to fit by any of the standard methods. In view of this it is all the more remarkable that the solution (19) gives as good a fit as it does with only four components, as will be seen on referring to my original publication, "The Progeny of a Population Element," p. 897, Fig. 4, already referred to.

Lastly, while the analogy is exact so long as we are dealing with industrial or organic aggregates maintained at a constant level, an essential difference arises when the case of a growing aggregate is considered. Organic growth takes place by what might be called "multiple replacement," that is, one individual in the course of life gives rise, on the average, to n individuals, where n may exceed unity. Analytically this finds expression in that

$$\int_0^\infty p(a)m(a)\,da > 1$$

and the fact is automatically taken care of in the solution (19) by the fact that in such a case the single real root $r > 0$.

Growth of industrial equipment, on the other hand, takes place by new units being installed *in addition to* replacement of disused units. The fundamental equations must be altered accordingly to take care of this case.

In conclusion I want to make a remark regarding the function of such analyses as the one here presented. In this connection I can do no better than to quote a sentence from Cournot:[30] "Those skilled in mathematical analysis know that

[30] A. Cournot, *Researches into the Mathematical Principles of the Theory of Wealth*, translated by N. Bacon, Macmillan Co., 1897, p. 3.

its object is not simply to calculate numbers, but that it is also employed to find the relations between magnitudes . . . ."

It is essentially in this sense that the analysis of a problem of industrial replacement is here offered. If we are merely interested in numbers, the direct arithmetical approach as practiced by Kurtz may be as good as any. But if an insight into the anatomy of the processes involved, and into their evolution from an initial condition to a final state is desired, then the setting up of the fundamental equations, and their solution in exponential series or in other suitable analytical form, and a concise expression of the relation between the distributions in time of successive generations, or orders of replacements, have greatly superior merit as compared with brute attacks by arithmetic without regard to mathematical form. Nor are the systematic relations (in terms of certain seminvariants) that have been shown to exist between the distribution of successive generations to be regarded merely as "short cuts" for their computation, though sometimes they may be found convenient in that way. Their real significance lies in that they serve to complete for us the analytical picture of the process of evolution of the system under consideration.

Metropolitan Life Insurance Company,
New York, N. Y.

BIBLIOGRAPHY

The following is a partial list of publications bearing on the general problem of self-renewing aggregates. The list is not the result of an exhaustive search, but merely of notes made in passing. Items marked with an asterisk are those referred to in the text. References are arranged essentially in chronological order.

1907 Lotka, A. J. "Relation between Birth Rates and Death Rates," *Science*, vol. 26, p. 21.

"Studies of the Mode of Growth of Material Aggregates," *American Journal of Science*, vol. 24, pp. 199, 375.

*1908 Hertz, P. "Die Bewegung eines Elektrons unter dem Einflusse einer stets gleich gerichteten Kraft," *Mathematische Annalen*, vol. 65, p. 1.

Herglotz, G. "Über die Integralgleichungen der Elektronentheorie," *Mathematische Annalen*, vol. 65, p. 87.

*1909 Herbelot, L. "Application d'un Théorème d'analyse à l'étude de la repartition par âge," *Bulletin Trimestriel de l'Institut des Actuaires Français*, vol. 19, p. 293.

*1911 Sharpe, F. R. and Lotka, A. J. "A Problem in Age Distribution," *Philosophical Magazine*, vol. 21, p. 435.

*1912 Risser, R. "Sur une application d'une équation fonctionnelle à un problème d'assurance," *Bulletin Trimestriel de l'Institut des Actuaires Français*, vol. 23, p. 298.

1913 Lotka, A. J. "A Natural Population Norm," *Journal of the Washington Academy of Sciences*, vol. 3, pp. 241, 289.

1918 Lotka, A. J. "The Relation between Birth Rate and Death Rate in a Normal Population and the Rational Basis of an Empirical Formula for the Mean Length of Life, given by William Farr," *Quarterly Publications of the American Statistical Association*, vol. 16, p. 121.

*1920 RISSER, R. "Sur une application de l'équation de Volterra au problème de la répartition par âge dans les milieux à effectif constant," *Comptes Rendus de l'Académie des Sciences*, p. 845.

BATEMAN, H. "An Integral Equation Occurring in a Mathematical Theory of Retail Trade," *Messenger of Mathematics*, vol. 49, p. 134.

1921 MOSER, C. "Über Gleichungen für eine sich erneuernde Gesellschaft mit Anwendung auf Sozialversicherungskassen," *Verhandlungen der Schweizerischen Naturforschenden Gesellschaft*, p. 9.

LOTKA, A. J. "A Simple Graphic Construction for Farr's Relation between Birth Rate, Death Rate and Mean Length of Life," *Journal of the American Statistical Association*, vol. 17, p. 998.

1922 LOTKA, A. J. "The Stability of the Normal Age Distribution," *Proceedings of the National Academy of Sciences*, vol. 8, p. 339.

1923 DOETSCH, G. "Die Integrodifferentialgleichungen vom Faltungstypus," *Mathematische Annalen*, vol. 89, p. 192.

1924 SCHOENBAUM, E. "Anwendung der Volterra'schen Integralgleichungen in der mathematischen Statistik" I, II, III, *Skandinavisk Aktuarietidskrift*, p. 241; 1925, p. 1.

FOCK, V. "Über eine Klasse von Integralgleichungen," *Mathematische Zeitschrift*, vol. 21, p. 161.

1925 LOTKA, A. J. *Elements of Physical Biology*, Williams & Wilkins Company, Baltimore, pp. 110, 117, 118 and 128.

DUBLIN, L. I. AND LOTKA, A. J. "On the True Rate of Natural Increase of a Population," *Journal of the American Statistical Association*, vol. 20, p. 569.

1926 LOTKA, A. J. "The Progressive Adjustment of Age Distribution to Fecundity," *Journal of the Washington Academy of Sciences*, vol. 16, p. 505.

NORTON, H. T. J. "Natural Selection and Mendelian Variation," *Proceedings of the London Mathematical Society*, Ser. 2, Part 1, vol. 28, p. 1.

MOSER, C. "Beiträge zur Darstellung von Vorgängen und des Beharrungszustandes bei einer sich erneuernden Gesamtheit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 21, p. 1.

MEIDELL, B. Über periodische und angenäherte Beharrungszustände," *Skandinavisk Aktuarietidskrift*, No. 3–4, p. 172.

WICKSELL, S. D. "Sveriges framtida befolkning under olika förutsättningar," *Ekonomisk Tidskrift*, p. 91.

CONNOR, L. R. "Fertility of Marriage and Population Growth," *Journal of the Royal Statistical Society*, Part III, vol. 89, p. 553.

DOETSCH, G. "Bemerkung zu der Arbeit von V. Fock: Über eine Klasse von Integralgleichungen," *Mathematische Zeitschrift*, vol. 24, p. 785.

1927 LOTKA, A. J. "The Size of American Families in the Eighteenth Century," *Journal of the American Statistical Association*, vol. 22, p. 154.

"The Actuarial Treatment of Official Birth Records," *Eugenics Review*, vol. 19, p. 257.

TRICOMI, F. "Risoluzione di un problema demografico," *Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 22.

INSOLERA, F. "Sul calcolo della probabilità annuale di morte in gruppi aperti di popolazione, "*Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 611.

"Su particulari equazioni di Volterra e loro applicazione finanziaria e demografica," *Rendiconti del R. Instituto Lombardo di Scienze e Lettere*, p. 247.

FRIEDLI, W. "Über die Stabilität der gegenseitigen Hilfskassen," *Zeitschrift für Schweizerische Statistik und Volkswirtschaft*, No. 3, vol. 63, p. 322.

HALDANE, J. B. S. "A Mathematical Theory of Natural and Artificial Selection," *Cambridge Philosophical Society Proceedings*, vol. 23, p. 607.

Frisch, R. "Sammenhengen mellem primaerinvestering og reinvestering," *Statsøkonomisk Tidskrift*, p. 117.

Doetsch, G. "Überblick über Gegenstand und Methode der Funktionalanalysis," *Jahresbericht der deutschen Mathematiker-Vereinigung*, vol. 36, p. 1.

Cantelli, F. P. "Intorno alla risoluzione di un problema demografico," *Atti della Reale Academia delle Scienze di Torino*, vol. 62, p. 449.

1928 Meidell, B. "Mathematical Statistics—On damping effects and approach to equilibrium in certain general phenomena," *Journal of the Washington Academy of Sciences*, No. 16, vol. 18, p. 437.

* Lotka, A. J. "The Progeny of a Population Element," *The American Journal of Hygiene*, No. 6, vol. 8, p. 875.

1929 Lotka, A. J. "The Spread of Generations," *Human Biology*, No. 3, vol. 1, p. 305.
"Biometric Functions in a Population Growing in Accordance with a Prescribed Law," *Proceedings of the National Academy of Sciences*, No. 10, vol. 15, p. 793.

Wyss, H. "Lage, Entwicklung und Beharrungszustand der eidgenössischen Versicherungskasse," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 24, p. 39.

1930 Jenzer, H. "Hauptgrössen der Witwenversicherung bei Einführung eines veränderlichen, exponentiellen Parameters für Witwensterblichkeit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 25, p. 203.

Wicksell, S. D. "Individens fruktsamhet och släktets förökning," *Ekonomisk Tidskrift*, Häft 2, p. 29.

Kurtz, E. B. *Life Expectancy of Physical Property*, Ronald Press Company, New York.

1931 Zwinggi, E. "Die Witwenversicherung als Teil der allgemeinen Alters—und Hinterlassenenversicherung," *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, Heft 26, p. 79.
"Das Problem der Erneuerung," *Festgabe Moser, Versicherungswissenschaftliche, juristische und mathematische Untersuchungen*, p. 127.

Husson, R. "Natalité et accroissement de la Population en France et à l'Étranger avant et après la guerre," *Bulletin de la Statistique générale de la France*, No. 2, vol. 20, p. 245.

Winfrey, R. and Kurtz, E. B. "Life Characteristics of Physical Property," *Iowa State College, Bulletin* 103, vol. 30, No. 3.

Zwinggi, E. "Zum Problem der Erneuerung," *Blätter für Versicherungs-Mathematik und verwandte Gebiete*, No. 1, vol. 31, p. 18.

Wyss, H. "Über die Prämiensysteme der Gruppenversicherung," *Festgabe Moser*, p. 175.

Wicksell, S. D. "Nuptiality, Fertility, and Reproductivity," *Skandinavisk Aktuarietidskrift*, p. 125.

Bonz, F. and Hilburg, F. "Die voraussichtliche Bevölkerungsentwicklung in Deutschland," *Zeitschrift für angewandte Mathematik und Mechanik*, No. 3, vol. 11, p. 237.

Goldstein, S. "Note on the Operational Solution of an Integral Equation," *Journal of the London Mathematical Society*, Part 4, vol. 6, p. 262.

Lotka, A. J. "The Structure of a Growing Population," Paper presented before the *Second General Assembly of the International Union for the Scientific Investigation of Population Problems*, London, 1931 and published in the *Proceedings* (Problems of Population) George Allen & Unwin, 1932. Also published in *Human Biology*, vol. 3, p. 459.

*1933 Lotka, A. J. "Industrial Replacement," *Skandinavisk Aktuarietidskrift*, 1933, p. 51.

"Applications de l'analyse au phénomène démographique," *Journal de la Société de Statistique de Paris*, vol. 74, p. 336.

Zwinggi, E. "Methodischer Beitrag zur Deckungs-Kapitalberechnung in der Sozialversicherung," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 28, p. 77.

v. Mises, R. Über die Vorausberechnung von Umfang und Altersschichtung der Bevölkerung Deutschlands," *Blätter für Versicherungs-Mathematik und verwandte Gebiete*, No. 10, vol. 2, p. 359.

1934 Maret, A. *Untersuchungen über diskontinuierlich sich erneuernde Gesamtheiten.* Inaugural Dissertation, Bern.

Wicksell, S. D. "Bidrag till den formella befolkningsteorien," *Statsokonomisk Tidsskrift*, Heft 1–2, p. 1.

*1935 Schulthess, H. *Volterra'sche Integralgleichungen in der Versicherungsmathematik.* Inaugural Dissertation der Philosophischen Fakultät II der Universität Bern.

1936 Lotka, A. J. "The Geographic Distribution of Intrinsic Natural Increase in the United States, and an Examination of the Relation between Several Measures of Net Reproductivity." *Journal of the American Statistical Association*, vol. 31, p. 273.

Lotka, A. J. "Modern Trends in Birth Rate," *The Annals of the American Academy of Political and Social Science*, vol. 188, November, p. 1.

Schulthess, H. "Über die Verwendung von Integralgleichungen zur Berechnung der mathematischen Reserve (Deckungskapital)," *Blätter für Versicherungs Mathematik und verwandte Gebiete*, vol. 3, p. 412.

Kermack, W. O. and McKendrik, A. G. "The Solution of Sets of Simultaneous Integral Equations Related to the Equation of Volterra," *Proceedings of the London Mathematical Society*, Ser. 2, vol. 41, p. 462; 472–477.

Kreis, H. "Stabilität einer sich jährlich erneuernden Gesamtheit," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 32, p. 17.

1937 Lotka, A. J. "A Historical Error Corrected," *Human Biology*, vol. 9, No. 1, p. 104.

"Population Analysis: A Theorem Regarding the Stable Age Distribution," *Journal of the Washington Academy of Sciences*, vol. 27, p. 299.

Schulthess, H. "Über das Erneuerungsproblem bei Verwendung eines analytischen Sterbegesetzes," *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, Heft 33, p. 69.

Hadwiger, H. "Zur Berechnung der Erneuerungsfunktion nach einer Formel von V. A. Kostitzin," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 34, p. 37.

Lotka, A. J. "Quelques résultats récents de l'analyse démographique," *Congrès International de la population*, Paris. I. Théorie générale de la population, p. 96.

Husson, R. "La démographie mathématique, Équations générales Propriétés générales des trajectoires démographiques," *Congrès International de la Population*, Paris. I. Théorie générale de la population, p. 108.

1938 Lotka, A. J. "Some Recent Results in Population Analysis," *Journal of the American Statistical Association*, vol. 33, p. 164.

*1938 Preinreich, G. A. D. "Annual Survey of Economic Theory: The Theory of Depreciation," *Econometrica*, vol. 6, p. 219.

Zwinggi, E. "Bemerkungen zum Erneuerungsproblem," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Heft 36, p. 69.

Hadwiger, H. "Über die Entwicklung der Erneuerungsfunktion in eine Potenzreihe," *Skandinavisk Aktuarietidskrift*, Häft 1–2, p. 31.

ON THE MATHEMATICS OF THE REPRESENTATIVE METHOD OF SAMPLING[1]

By Allen T. Craig

1. **Introduction.** This paper is designed to present certain topics in mathematical statistics which find application in some of the problems that arise in what has been termed the representative method of sampling.

For descriptive purposes, it seems convenient to consider two aspects of the representative method. The first of these may be called the method of *purposive selection*. This method can be roughly characterized by saying that it is the method employed when the samples are chosen in such a way that each sample will possess one or more characters, say certain averages, which are identical with the corresponding characters in the population from which the samples are drawn. The mathematical conditions which underlie this method are rather stringent, and both theoretical and practical investigations seem to have proved that in general no great amount of confidence can be placed in the results obtained.

The second aspect of the representative method has been styled the method of *random sampling*. This method can take either of two forms which we may call the method of *unrestricted random sampling* and *stratified random sampling*. The first of these is the classical method of procedure. That is, a sample is drawn at random from a given population and on the basis of these data inferences are made concerning the nature of the population. On the other hand, when the method of stratified random sampling is used, the population is first separated into a large number of parts, called strata, and the sample consists of an equally large number of "partial samples," each partial sample being drawn from a different stratum. It appears, both from theoretical and practical results, that this method of stratified random sampling enjoys many advantages not shared by the other methods.

We now turn to the main purpose of this paper, namely that of enumerating some of the theorems and methods of mathematical statistics which serve useful purposes in this theory. Discussion of how these theorems find application in the method itself has been reserved for other participants on this program.

2. **Estimates.** From our preliminary remarks, it is apparent that the representative method is much concerned with the problem of estimating certain

---

[1] Presented, at the invitation of the program committee, to a joint session of the Institute of Mathematical Statistics and the American Statistical Association on December 29, 1938.

unknown *parameters* of a statistical population. On this account, we first consider the problem of estimates.

Consider a population with arithmetic mean m and standard deviation σ. Let $x_1, x_2, \cdots, x_n$, be n independent items drawn from this population and let $c_1, c_2, \cdots, c_n$ be any finite real constants, not all zero to avoid the trivial case. Write $y = c_1x_1 + c_2x_2 + \cdots + c_nx_n$. Then the expected or arithmetic mean value of y is

$$\bar{y} = E(y) = m(c_1 + c_2 + \cdots + c_n),$$

and the variance of y is

$$\sigma_y^2 = E\{(y - \bar{y})^2\} = \sigma^2(c_1^2 + \cdots + c_n^2).$$

Suppose we inquire into the probability that y will have a value which is within a preassigned ϵ of its expected value. To this end, let C be the numerical value of the numerically greatest of the set $c_1, \cdots, c_n$, so that $\sigma_y^2 \leq n\sigma^2C^2$. Then by Tchebycheff's inequality p, the probability that $|y - \bar{y}| < \epsilon$, where ϵ is an arbitrarily small positive number, is such that

$$p \geq 1 - \frac{\sigma_y^2}{\epsilon^2},$$

or

$$p \geq 1 - \frac{n\sigma^2C^2}{\epsilon^2}.$$

In general, this inequality will have little interest. But if C is of the form $M/n^{\frac{1+\delta}{2}}$, M independent of n, $\delta > 0$, then $p \geq 1 - \frac{\sigma^2M^2}{n^\delta\epsilon^2}$ and by increasing n the right member can be made as near to one as we please. This means then that if we have a population with a finite variance and if we construct a linear function of the observations with coefficients of the nature indicated, we can, by increasing the size of the sample, make the probability approach one that the linear function will have a value arbitrarily close to its expected value.[2]

Now suppose that instead of constructing an arbitrary linear function we attempt to construct a function which will be an estimate of some particular parameter of the population. If the estimate is to be most serviceable, we should like to be able, by governing the size of the sample, to be as certain as we like that the estimate will have a value arbitrarily near that of the parameter. The preceding discussion shows that we can best achieve this by requiring that the expected value of the estimate be equal to the parameter sought. An estimate such as that just described is frequently called an *unbiased* estimate. The use of such estimates in statistical problems makes it possible to avoid systematic errors in estimating parameters. In general, unique unbiased estimates of a parameter do not exist. For example, the arithmetic mean m of

[2] Under these conditions, the function of the observations is said to converge stochastically to its expected value.

the population can be estimated from the sample $x_1, \cdots, x_n$ by any one of a large number of unbiased estimates such as $(x_1 + x_2 + \cdots + x_n)/n$, $(x_1 + x_n)/2$, x_4, and so on without limit. Thus it becomes necessary to make a choice of the unbiased estimate to be used. An appropriate criterion is that the unbiased estimate whose distribution has the smallest variance is the best to use. The reason for this can be seen by examining the preceding formula $p \geq 1 - \frac{\sigma_y^2}{\epsilon^2}$. For if y_1 and y_2 are two unbiased estimates of the same parameter and if $\sigma_{y_1}^2 < \sigma_{y_2}^2$, then in $p_1 \geq 1 - \frac{\sigma_{y_1}^2}{\epsilon^2}$ and $p_2 \geq 1 - \frac{\sigma_{y_2}^2}{\epsilon^2}$ we see that $1 - \frac{\sigma_{y_1}^2}{\epsilon^2}$ is more nearly equal to one than is $1 - \frac{\sigma_{y_2}^2}{\epsilon^2}$. Because of this fact we prefer, at least in most problems, to use y_1 rather than y_2 as an estimate of the unknown parameter. An unbiased estimate whose sampling variance is a minimum is sometimes called a *best estimate*.[3] It should not be inferred that the word "best" has any implications other than those stated explicitly in the definition.

The question very naturally arises as to whether we can determine these best estimates in particular cases. In general we can not determine them, but under certain conditions we can find best estimates if we are dealing with linear functions of the observations. A method and the conditions are set forth in an important theorem due to Markoff. We now consider his method.

3. **Markoff's Method.** Let there be given n statistical populations with arithmetic means $m_1, m_2, \cdots, m_n$ and standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_n$ respectively. We assume that no correlation exists between any of the populations. Furthermore, suppose that each of the n arithmetic means can be expressed linearly in terms of k unknown, but unique, parameters, say $z_1, z_2, \cdots, z_k$. Thus

$$
\begin{aligned}
m_1 &= a_{11}z_1 + a_{12}z_2 + \cdots + a_{1k}z_k \\
m_2 &= a_{21}z_1 + a_{22}z_2 + \cdots + a_{2k}z_k \\
&\vdots \\
m_n &= a_{n1}z_1 + a_{n2}z_2 + \cdots + a_{nk}z_k,
\end{aligned}
\tag{1}
$$

where the a's are known constants. Likewise, let T be a parameter which is expressible linearly in terms of the same k unknown parameters, say $T = b_1z_1 + b_2z_2 + \cdots + b_kz_k$, where the b's are given constants. We draw a sample of n independent items, $x_1, x_2, \cdots, x_n$, in which one item is drawn from each

[3] An estimate of a parameter which converges stochastically (cf. footnote (2)) to that parameter is called a *consistent* estimate of the parameter. If a consistent estimate has a distribution which is normal for large samples and if the variance of that distribution is smaller than the variance of any other consistent estimate which also has a normal distribution for large samples, then the estimate is called *efficient*. It should be observed that our definition of best estimate requires an unbiased estimate, whereas consistent and efficient estimates may be biased.

of the n populations. On the basis of this sample we seek to determine a set of numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$ such that $T' = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n$ is the best estimate of T.

Before attempting to find the solution, if one exists, let us first examine the mathematical implications of the problem. In the first place, in order that parameters $z_1, \cdots, z_k$ may exist, it is necessary and sufficient that the matrices A and B, where

$$A = \begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{Vmatrix} \text{ and } B = \begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1k} & m_1 \\ a_{21} & a_{22} & \cdots & a_{2k} & m_2 \\ \vdots & & & & \\ a_{n1} & a_{n2} & \cdots & a_{nk} & m_n \end{Vmatrix}$$

have the same rank. Thus we require that A and B have the common rank R. This being satisfied, we note further that if $k > n$, there will be infinitely many values of the z's which will satisfy the equations (1). Thus we require in addition that $k \leq n$. Finally, we note that if the common rank R is less than k, there will be infinitely many values of the z's which will satisfy the system (1). Hence we must have $R = k \leq n$.

We now turn to a consideration of the solution of the problem. Whatever the values of the λ's, we have for the mean value and the variance of T'

$$\begin{aligned} E(T') &= \lambda_1 m_1 + \cdots + \lambda_n m_n \\ &= \lambda_1 \Sigma a_{1j} z_j + \cdots + \lambda_n \Sigma a_{nj} z_j, \end{aligned}$$

and

$$\sigma_{T'}^2 = \lambda_1^2 \sigma_1^2 + \cdots + \lambda_n^2 \sigma_n^2,$$

respectively. Since $E(T')$ must equal T as a part of the condition for a best estimate, then

$$\lambda_1 \Sigma a_{1j} z_j + \cdots + \lambda_n \Sigma a_{nj} z_j = b_1 z_1 + \cdots + b_k z_k$$

identically in the z's. That is, the coefficients of $z_1, \cdots, z_k$ in the left member must equal the corresponding coefficient in the right member. Accordingly,

$$(2) \qquad \begin{aligned} a_{11}\lambda_1 + a_{21}\lambda_2 + \cdots + a_{n1}\lambda_n &= b_1 \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \cdots + a_{n2}\lambda_n &= b_2 \\ &\vdots \\ a_{1k}\lambda_1 + a_{2k}\lambda_1 + \cdots + a_{nk}\lambda_n &= b_k . \end{aligned}$$

If these equations are to have solutions for $\lambda_1, \cdots, \lambda_n$, we must make the additional assumption that the matrix C, where

$$C = \begin{Vmatrix} a_{11} & a_{21} & \cdots & a_{n1} & b_1 \\ a_{12} & a_{22} & \cdots & a_{n2} & b_2 \\ \vdots & & & & \\ a_{1k} & a_{2k} & \cdots & a_{nk} & b_k \end{Vmatrix}$$

has the same rank as the matrix of the coefficients, namely R. If this condition is satisfied we can write equations (2) in the form

$$\begin{aligned} a_{11}\lambda_1 + \cdots + a_{k1}\lambda_k &= b_1 - a_{k+1,1}\lambda_{k+1} - \cdots - a_{n1}\lambda_n \\ &\vdots \\ a_{1k}\lambda_1 + \cdots + a_{kk}\lambda_k &= b_k - a_{k+1,k}\lambda_{k+1} - \cdots - a_{nk}\lambda_n \end{aligned} \tag{3}$$

and solve for $\lambda_1, \cdots, \lambda_k$ in terms of the a's, the b's, and $\lambda_{k+1}, \cdots, \lambda_n$. Here, without any essential loss of generality, we take the non-vanishing k-rowed determinant to be that of the coefficients of $\lambda_1, \cdots, \lambda_k$ in equations (2). Thus for arbitrarily assigned values of $\lambda_{k+1}, \cdots, \lambda_n$, we can compute the values of $\lambda_1, \cdots, \lambda_k$ and these n values of the λ's will give us a T' which is an unbiased estimate of T. That there will be, in general, an unlimited number of sets of values of the λ's is in keeping with our previous observation that unique unbiased estimates usually do not exist.

The next part of the problem will consist in determining which, if any, of the above sets of λ's will make $\sigma_{T'}^2$ a minimum. We recall that $\sigma_{T'}^2 = \lambda_1^2\sigma_1^2 + \cdots + \lambda_n^2\sigma_n^2$. In $\sigma_{T'}^2$ let us replace $\lambda_1, \cdots, \lambda_k$ by their values (in terms of $\lambda_{k+1}, \cdots, \lambda_n$) which we obtained by solving the system (3). Then $\sigma_{T'}^2$ will be expressed in terms of $\sigma_1, \cdots, \sigma_n$, the a's, the b's, and $\lambda_{k+1}, \cdots, \lambda_n$. We next take the partial derivative of $\sigma_{T'}^2$ with respect to each of $\lambda_{k+1}, \cdots, \lambda_n$. On equating these partial derivatives to zero we will have a system of $n - k$ linear equations in the $n - k$ unknowns $\lambda_{k+1}, \cdots, \lambda_n$. If these equations yield unique values for $\lambda_{k+1}, \cdots, \lambda_n$, they will in turn determine unique values of $\lambda_1, \cdots, \lambda_k$. This gives us a unique set of λ's such that at one and the same time

$$E(T') = T \text{ and } \sigma_{T'}^2 \text{ is a minimum.}$$

The procedure which we have just outlined is most tedious to carry out in a particular case. Because of the insight of Markoff, a much better scheme is available for finding the best estimate of T. Consider the function of $z_1, \cdots, z_k$,

$$\begin{aligned} F(z_1, \cdots, z_k) &= \sum \left(\frac{x_j - m_j}{\sigma_j}\right)^2 \\ &= \sum \left(\frac{x_j - a_{j1}z_1 - \cdots - a_{jk}z_k}{\sigma_j}\right)^2. \end{aligned}$$

Evaluate $\dfrac{\partial F}{\partial z_1}, \cdots, \dfrac{\partial F}{\partial z_k}$ and equate these partial derivatives to zero. This yields the following system of k linear equations in the k unknowns $z_1, \cdots, z_k$.

$$\begin{aligned} z_1 \sum \frac{a_{j1}^2}{\sigma_j^2} + \cdots + z_k \sum \frac{a_{j1}a_{jk}}{\sigma_j^2} &= \sum \frac{a_{j1}x_j}{\sigma_j^2} \\ &\vdots \\ z_1 \sum \frac{a_{j1}a_{jk}}{\sigma_j^2} + \cdots + z_k \sum \frac{a_{jk}^2}{\sigma_j^2} &= \sum \frac{a_{jk}x_j}{\sigma_j^2}. \end{aligned} \tag{4}$$

If the system (4) yields unique values for the z's, these values, when substituted in T, yield exactly the same estimate of T as was found by substituting for the λ's in T'.

Perhaps an illustration will make this clearer. Suppose we have $n = 2$ populations and that the means m_1 and m_2 are expressible linearly in terms of $k = 1$ parameter z_1. Our equations (1) become

$$m_1 = a_{11}z_1 \qquad (1')$$
$$m_2 = a_{21}z_1 .$$

Similarly, we have $T = b_1z_1$ and $T' = \lambda_1x_1 + \lambda_2x_2$. We first determine the λ's such that T' is the best estimate of T. In accordance with the preceding steps, equations (2) become

$$a_{11}\lambda_1 + a_{21}\lambda_2 = b_1, \qquad (2')$$

and the system (3) becomes

$$\lambda_1 = \frac{b_1 - a_{21}\lambda_2}{a_{11}}, \qquad a_{11} \neq 0. \qquad (3')$$

Then

$$\begin{aligned}\sigma_{T'}^2 &= \lambda_1^2\sigma_1^2 + \lambda_2^2\sigma_2^2 \\ &= \left(\frac{b_1 - a_{21}\lambda_2}{a_{11}}\right)^2 \sigma_1^2 + \lambda_2\sigma_2^2,\end{aligned}$$

because of (3′). Thus

$$\frac{\partial\sigma_{T'}^2}{\partial\lambda_2} = \frac{-2a_{21}(b_1 - a_{21}\lambda_2)\sigma_1^2}{a_{11}^2} + 2\lambda_2\sigma_2^2,$$

and for a minimum $\sigma_{T'}^2$ we write $\frac{\partial\sigma_{T'}^2}{\partial\lambda_2} = 0$ so that

$$\lambda_2 = \frac{a_{21}b_1\sigma_1^2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

Since

$$\lambda_1 = (b_1 - a_{21}\lambda_2)/a_{11} ,$$

then

$$\lambda_1 = \frac{b_1a_{11}\sigma_2^2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

Our best estimate of T is found from T' and it is

$$T' = \frac{b_1a_{11}\sigma_2^2x_1 + b_1a_{21}\sigma_1^2x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}.$$

By Markoff's method we would form the function

$$F(z_1) = \sum \left(\frac{x_j - a_{j1} z_1}{\sigma_j}\right)^2 = \left(\frac{x_1 - a_{11} z_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - a_{21} z_1}{\sigma_2}\right)^2.$$

The system (4) reduces to merely

$$z_1 = \frac{a_{11}\sigma_2^2 x_1 + a_{21}\sigma_1^2 x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2}. \tag{4'}$$

We substitute this value of z_1 in $T = b_1 z_1$ and obtain

$$T = \frac{b_1 a_{11}\sigma_2^2 x_1 + b_1 a_{21}\sigma_1^2 x_2}{a_{11}^2\sigma_2^2 + a_{21}^2\sigma_1^2},$$

which is the estimated value T' above.

4. **Neyman's modification of Markoff's Method.** We are indebted to Neyman for a modification and adaptation of the Markoff method so as to make the method applicable to some of the problems of stratified random sampling. One of his examples will best illustrate the method.

Suppose that a given population is divided into n strata. Let the jth stratum contain M_j items and let these items be $u_{j1}, u_{j2}, \cdots, u_{jM_j}$. The mean and the variance of this stratum are then

$$\bar{u}_j = \frac{1}{M_j}\sum_k u_{jk} \quad \text{and} \quad \sigma_j^2 = \frac{1}{M_j}\sum_k (u_{jk} - \bar{u}_j)^2.$$

Let T be the parameter $T = M_1\bar{u}_1 + M_2\bar{u}_2 + \cdots + M_n\bar{u}_n$, so that $\dfrac{T}{M_1 + \cdots + M_n}$, the mean of the population, is expressed as a linear function of the means of the n strata. We draw at random a sample of N items, the sample consisting of n partial samples, one partial sample being drawn from each of the n strata. Suppose there are n_1 items in the partial sample from the first stratum, n_2 from the second, and so on. Thus $n_1 + n_2 + \cdots + n_n = N$ and the entire sample consists of the n partial samples

$$\begin{array}{l} x_{11}, x_{12}, \cdots, x_{1n_1} \\ x_{21}, x_{22}, \cdots, x_{2n_2} \\ x_{n1}, x_{n2}, \cdots, x_{nn_n}. \end{array}$$

From these N data we propose constructing an estimate

$$T' = \lambda_{11}x_{11} + \cdots + \lambda_{1n_1}x_{1n_1} + \cdots + \lambda_{n1}x_{n1} + \cdots + \lambda_{nn_n}x_{nn_n}$$

which will be the best estimate of T. Now the expected value of T' is

$$\begin{aligned} E[T'] = E\left[\sum_j \sum_k \lambda_{jk} x_{jk}\right] &= \sum_j \sum_k \lambda_{jk} E(x_{jk}) \\ &= \sum_j \sum_k \lambda_{jk} \bar{u}_j \\ &= \sum_j \bar{u}_j \sum_k \lambda_{jk}, \end{aligned}$$

which, by hypothesis, must equal T. Thus

$$\sum_1^n \bar{u}_j \sum_1^{n_j} \lambda_{jk} = \sum_1^n M_j \bar{u}_j$$

identically in the $\bar{u}$'s. Hence $\Sigma \bar{u}_j (M_j - \Sigma \lambda_{jk}) = 0$ which requires that the coefficients of $\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_n$ must be zero. That is

$$\sum_1^{n_1} \lambda_{1k} = M_1$$

$$\vdots$$

$$\sum_1^{n_n} \lambda_{nk} = M_n.$$

Of course there are infinitely many λ's which will satisfy these equations. But we can eliminate all but one set by imposing the condition that $\sigma^2_{T'}$ shall be a minimum. The algebra of mathematical expectation can be used to show that

$$\sigma^2_{T'} = \sum_1^n \sigma^2_j \left[\frac{M_j n_j - n_j^2}{M_j - 1} \left(\frac{1}{n_j} \sum \lambda_{jk} \right)^2 + \frac{M_j}{M_j - 1} \sum \left(\lambda_{jk} - \frac{1}{n_j} \sum \lambda_{jk} \right)^2 \right]$$

which will be a minimum when $\sum \left(\lambda_{jk} - \frac{1}{n_j} \sum \lambda_{jk} \right)^2 = 0, j = 1, 2, \cdots, n$. Since this is a sum of real squares, each term in the sum must be zero. Thus, $\lambda_{jk} = \frac{1}{n_j} \sum \lambda_{jk}$. Since $\sum \lambda_{jk}$ must equal M_j in order that $E(T') = T$, then $\lambda_{jk} = \frac{M_j}{n_j}$ which uniquely determines the λ's and hence our best estimate of T'.

It is important to observe that Neyman's adaptation does not assume that the various strata are uncorrelated nor that there are necessarily replacements after each drawing in taking the sample.

5. **Estimation of Ratios.** In certain problems in representative sampling it may be necessary to estimate both the numerator and the denominator of a fraction, say T/U. If T' and U' are linear estimates of T and U then for large samples both T' and U' will be approximately normally distributed in most cases. Further, if T' and U' are correlated, they will usually be approximately normally correlated. Geary has proved that if we write

$$V = \frac{b + T'}{a + U'},$$

where a and b are constants and U' and T' are measured from their expected values, then

$$t = \frac{aV - b}{\sqrt{V^2 \sigma^2_{U'} - 2rV\sigma_{T'}\sigma_{U'} + \sigma^2_{T'}}}$$

is approximately normally distributed with mean zero and unit variance provided $a \geq 3\sigma_{U'}$. Here r is the correlation coefficient between T' and U'. For

large samples this provides a convenient method of testing the significance of the difference between an observed and a hypothetical ratio of two linear estimates.

6. Fiducial Inference. After an estimate of a parameter has been made, it is usually desirable to make some inference about the true value of the parameter. For many years the concept of probable error was used in this connection. But the use of the probable error involves the assumption that all values of the unknown parameter are equally likely. This assumption is questionable and efforts to avoid making the assumption have led to a theory called *fiducial inference*. This method of statistical inference has broad implications but limitations on our time do not permit our discussing the topic. At the close of this paper, we give certain references to the subject, including some of an expository nature.

7. Conclusion. As stated in the introduction, this paper purports to give an exposition of some of the topics in mathematical statistics which find application in the representative method of sampling. Necessarily considerable selection of material had to be made. We believe, however, that the problem of the best estimate and an appropriate method of obtaining such an estimate are fundamental, and we hope that our exposition has helped to make clear these concepts of mathematical statistics which have proved so useful in the representative method.

REFERENCES

[1] J. Neyman, "On the two different aspects of the representative method," *Journal of the Royal Statistical Society*, vol. 97 (1934) pp. 558–625.

[2] J. Neyman, *Lectures and conferences on mathematical statistics, Graduate School of the U. S. Department of Agriculture*, (1937) pp. 89–105.

[3] J. Neyman, "Contributions to the theory of sampling human populations," *Journal of the American Statistical Association*, vol. 33 (1938) pp. 101–116.

[4] S. S. Wilks, *Statistical inference*, Edwards Brothers (1936–37) pp. 8–11, p. 70, pp. 71–86.

[5] R. C. Geary, "The frequency distribution of the quotient of two normal variates," *Journal of the Royal Statistical Society*, vol. 93 (1930) pp. 442–446.

[6] R. A. Fisher, "Inverse probability," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930) p. 528.

[7] H. L. Rietz, "A recent advance in statistical inference," *American Mathematical Monthly*, vol. 45 (1938) pp. 149–158.

The University of Iowa.

ON A NEW CLASS OF "CONTAGIOUS" DISTRIBUTIONS, APPLICABLE IN ENTOMOLOGY AND BACTERIOLOGY

BY J. NEYMAN

CONTENTS

PAGE
1. Introduction 35
2. Distribution of larvae in experimental plots 36
3. Particular classes of the limiting distribution of X 40
4. Certain general properties of the distributions deduced 43
5. Contagious distribution of type A depending on two parameters 45
6. Contagious distribution of type A depending on three parameters 48
7. Contagious distributions of types B and C 53
8. Illustrative examples and concluding remarks 54
9. References 57

1. **Introduction.** There are a number of fields in which experimental data cannot be treated with any success by means of the usual "Student's" test and—very probably—by means of the more general analysis of variance z-test of Fisher. It is known in fact [1] that the t-test, as applied to two samples, is only valid when the populations from which the samples are drawn have equal variances. As the z-test is of a nature similar to the t-test, with the difference that it is applied to detect differentiation in means of more than two populations, a similar conclusion seems very likely. Thus, whenever we have to compare means of populations with distinctly different variances, we have to look for some new tests. It may be useful to mention at once two instances in which the situation mentioned actually arises.

As a first instance we may quote certain entomological experiments. Suppose it is desired to test the efficiency of several treatments intended to destroy certain larvae on a field. The experiments are arranged in the usual way. The treatments compared are applied to particular plots with several replications and then the plots (or smaller parts of them) are inspected and all the surviving larvae are counted. Thus the observations represent the numbers of surviving larvae in several equal areas. It happens frequently that, while there is room for doubt as to whether there is any significant difference between the average number of survivors corresponding to particular treatments, there is no doubt whatever that the variability of the observations differs from treatment to treatment.

We have another similar case in bacteriology. The experiments I have in mind consist in determining the bacterial density by the so called "plating method." This consists in taking a number of samples of the analyzed liquid

and in spreading them separately on Petri plates. After a suitable period of time a number of colonies appear on the plates and their numbers represent the observational figures. I am informed that the variability of such observations does not depend very much on the technique of mixing the liquid and of taking the samples—when this technique is on a proper level—but does considerably depend on the kind and on the number of bacteria present in the liquid.

The above examples justify an effort to find some new and more appropriate test. The first step in this direction must consist in an analysis of the machinery behind the observable distributions and in deducing their analytical form. Once this problem is solved and repeated comparisons show a satisfactory agreement between the theory and the observation, we may proceed to the next step and deduce the appropriate tests.

The purpose of the present paper consists in deducing a family of distributions which provide a reasonably good fit in several cases in which they have been tested. It may be hoped that they will prove satisfactory also in many cases in the future.

2. **Distribution of larvae in experimental plots.** When the problem of the distribution of larvae in experimental plots first arose, attempts were made to fit the Poisson Law of frequency. These attempts, however, failed almost invariably with the characteristic feature that, as compared with the Poisson Law, there were too many empty plots and too few plots with only one larva. A similar circumstance is frequently, though not so regularly, observed in counts of microorganisms in single squares of a haemacytometer. These facts suggest that the distributions considered belong to a class which Pólya [3] proposed to call "contagious": the presence of one larva within an experimental plot increases the chance of there being some more larvae. And it is not difficult to see the cause of this dependence. Larvae are hatched from eggs which are being laid in so-called "masses." After being hatched they begin to travel in search of food. Their movements are slow and therefore, whenever in a given plot we find a larva, this means that the mass of eggs, from which it was hatched, must have been laid somewhere near, and this in turn means that we are likely to find in the same plot some more larvae from the same litter. Of course, there may be also others coming from other litters, too.

A similar explanation may apply also to microorganisms counted in single squares of a haemacytometer or to colonies on parallel plates. However, here the situation does not seem as clear as in the case of larvae. As far as the haemacytometer counts are concerned, also another cause of contagiousness may be suggested. Witnessing once the process of preparation of the experiment, I noticed that, immediately after the drop of liquid was deposited into the chamber of the haemacytometer and for some time after, the positions of cells seen under the microscope were not fixed. Some of them seemed to lie on the bottom and the others were floating downwards in an irregular movement. Trying to follow the movements of particular cells I had the impression

that they were slightly attracted by the cells already stationary or semi-stationary on the bottom of the chamber. If this impression of mine is justified, then the attraction of the floating cells by those already on the bottom could explain the contagiousness of the resulting distribution. It is known, however, that this contagiousness is always rather small and that frequently the distribution of cells in the squares of the haemacytometer does follow the Poisson Law very closely.

Owing to the fact that the cause of the contagiousness of the distribution of larvae in experimental plots is clear, we shall deal primarily with the distribution of larvae. Consequently, if the theoretical distributions that we shall deduce fit the empirical ones, we shall be more or less justified in assuming that we guessed the essential features of the actual machinery of movements of the larvae. On the other hand, if the same theoretical distributions appear also to fit satisfactorily the empirical counts of bacteria then in respect of these applications it will be safer to consider that we were lucky enough to find a sufficiently flexible interpolation formula.

After these preliminaries we may proceed to a more accurate specification of the conditions of the problem considered. The experimental plot in which the larvae are counted will be denoted by P. We shall make no restriction as to the shape of this plot, but we shall assume that its area, which we shall take as unity, is small compared with that of the experimental field, F. The latter will be assumed to possess M units of area. We shall further assume that the moths laying eggs on the field F select spots for this purpose in a purely random manner. This presupposes that the experimental field is uniform in many relevant respects, e.g. is sown in all its parts by the same kind of plant, etc. Denoting by ξ and η the coördinates of the mass of eggs laid by some particular moth on the field F, we shall treat them as random variables with the elementary probability law

$$p(\xi, \eta) = \frac{1}{M} \tag{1}$$

everywhere within F and zero elsewhere. After the larvae are hatched from the eggs there will be some mortality among them. Let us denote by n the number of larvae hatched from the same mass of eggs, surviving at the moment when the counts are made. We shall treat n as a random variable and denote by $p(n)$ its probability law. At the present moment the writer has no information as to what may be the nature of the function $p(n)$. Consequently it will remain in our calculations in its general form and, wishing to obtain some formulae for immediate calculations, we shall have to substitute for $p(n)$ hypothetical formulae which, on intuitive grounds, may seem plausible. If the larvae counted are all more or less of the same age, there is a possibility that $p(n)$ does not differ very much from the Poisson Law, but this point might be verified experimentally and we shall not insist on its being necessarily true.

Consider now a single larva, survivor at the moment of observation, which

was hatched out at a point with coördinates ξ and η. Denote by x and y the coördinates of this larva at the moment of counts. We shall consider x and y as random variables. It is obvious that the probability law of x and y must depend on the values of ξ and η. We shall assume that the dependence is of a particular character; namely, that the probability law of x and y given ξ and η is a function of the differences $x - \xi$ and $y - \eta$. We shall denote it by $f(x - \xi, y - \eta)$.

There is very little that we may consider as known about the function $f(x - \xi, y - \eta)$. It may be treated as describing the habits of travelling of the larvae. There are some indications that there are certain directions in which the larvae tend to travel rather than in others, but they are too vague to be taken into consideration. Only one thing is certain: during the period of time between the birth of the larvae and the moment that the counts are made the larvae are able to travel only at some limited distance. Consequently we shall assume that for sufficiently large values of $| x - \xi |$ and $| y - \eta |$ the function $f(x - \xi, y - \eta)$ is identically zero. Otherwise we shall not make any further assumption concerning $f(x - \xi, y - \eta)$, and it will remain arbitrary in our calculations until we reach the final general formula.

While abstaining from making arbitrary assumptions concerning the habits of single larvae, we shall make one concerning the habits of several of them. This assumption, however, seems to be very plausible. We shall assume that the larvae have no social instincts, so that the random variables x and y corresponding to one larva are independent from those corresponding to any other —that is to say, apart from the possible dependence on the same pair of ξ and η.

Denote by N the total number of masses of eggs laid on the field F and let k_i be the number of larvae hatched from the i-th mass of eggs, surviving at the moment of observation and present within some particular experimental plot P. Finally let

$$X = \sum_{i=1}^{N} k_i \tag{2}$$

be the total number of larvae to be found within this plot. Our purpose will be to use the above hypotheses in order to determine the probability law of X. In doing so we shall first find that of any of the k_i's. Obviously, when considering just one variable k_i, it would be useless to retain the subscript i, so that below we shall write simply k to denote the number of living larvae, to be found within P, all of which were hatched from the same mass of eggs, situated at some point (ξ, η).

Let us first write the expression for the probability that one particular larva of that group will be found within P. This probability will be a function of ξ and η only, say

$$P(\xi, \eta) = \int\int_P f(x - \xi, y - \eta)\, dx\, dy. \tag{3}$$

Given that the number of survivors of the mass of eggs of the point (ξ, η) is n, the probability that exactly k of them will be found within P will be represented by the binomial formula, say

$$P\{k \mid n, \xi, \eta\} = \frac{n!}{k!\,(n-k)!} P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}. \tag{4}$$

It will be noticed that in writing this formula we use the hypothesis that the larvae have no social instincts.

Multiplying (4) by the probability law of ξ and η, and integrating with respect to those variables over the whole field F, we shall obtain the probability, $P\{k \mid n\}$ that out of the n survivors of a mass of eggs, laid anywhere within F, exactly k larvae will be found within P:

$$P\{k \mid n\} = \frac{n!}{k!\,(n-k)!} \frac{1}{M} \int \int_F P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta. \tag{5}$$

Multiplying this result by $p(n)$ and summing for all values of n, we shall obtain the absolute probability of k having any specified value.

However, before doing so, we must use the hypothesis about the function $f(x - \xi, y - \eta)$ to deduce certain consequences concerning the integral in (5).

Originally we did not make any assumption as to the origin of coördinates on the field F. It will be now convenient to assume that it is located somewhere within the experimental plot P, for example in its center or in any other easily specified point. Owing to the particular property of the function $f(x - \xi, y - \eta)$ it will now follow that, for sufficiently large values of ξ and η, the probability $P(\xi, \eta)$ will be equal to zero. Let us denote by A the part of the experimental field where $P(\xi, \eta) > 0$. Obviously A denotes the set of points, a, in F such that, if a mass of eggs is laid in one of them, the distance of a from the plot P is not too large for the larvae hatched in a to reach the plot P before the moment of observation. Obviously also the plot P is included in A. Consequently the area of A, to be denoted by the same letter A, must be greater than unity. Owing to the lack of any precise knowledge of the nature of the function $f(x - \xi, y - \eta)$ it is impossible to say anything about the shape of A.

Let us now turn to the integral in (5). The function under this integral changes its form according to whether the point (ξ, η) is within or without A. If $k = 0$, then the integral in (5) reduces to

$$\int \int_F (1 - P(\xi, \eta))^n\, d\xi\, d\eta = M - A + \int \int_A (1 - P(\xi, \eta))^n\, d\xi\, d\eta. \tag{6}$$

If however $k > 0$, then

$$\int \int_F P^k(\xi, \eta)\,(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta = \int \int_A P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}\, d\xi\, d\eta. \tag{7}$$

Now we can write

$$P\{k\} = \sum_{n \geq 0} p(n) P\{k \mid n\}, \tag{8}$$

which gives in particular

$$(9) \qquad P\{k = 0\} = 1 - \frac{A}{M} + \frac{1}{M} \int \int_A \sum_{n \geq 0} (1 - P(\xi, \eta))^n p(n)\, d\xi\, d\eta$$

and for $k > 0$

$$(10) \qquad P\{k\} = \frac{1}{M} \int \int_A \sum_{n \geq 0} \frac{n!}{k!\,(n-k)!} P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k} p(n)\, d\xi\, d\eta.$$

This is the general form of the probability law of k, which involves two unspecified functions $p(n)$ and $P(\xi, \eta)$. We shall not analyze it but proceed to the calculation of the characteristic function $\phi_k(t)$ of k, which will then be used to calculate that of X. We have

$$(11) \qquad \phi_k(t) = \sum_{k \geq 0} e^{itk} P\{k\}$$

or, using (9) and (10), and after easy transformations

$$(12) \qquad \phi_k(t) = 1 - \frac{A}{M}\left(1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right).$$

Owing to the assumption that the larvae have no social instincts all the variables $k_1, k_2, \cdots k_N$ in (2) must be considered as mutually independent. As the characteristic function of any of them has the same form (12), the characteristic function, $\phi_X(t)$, of their sum, X, will be represented by the Nth power of the expression (12). Denoting by m the average number of masses of eggs per unit of area of the field F, so that $N = Mm$, we shall have

$$(13) \qquad \begin{aligned} \phi_X(t) &= \phi_k^N(t) \\ &= \left\{1 - \frac{A}{M}\left(1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right)\right\}^{Mm}. \end{aligned}$$

This will be the characteristic function of X for any value of M. If it is desired to put into effect the assumption that "M is large", we shall have to consider the limit of (13) for $M \to \infty$. This will be denoted by $\phi(t)$ and we shall have

$$(14) \qquad \phi(t) = \exp\left\{- Am\left(1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n)(P(\xi, \eta)e^{it} + 1 - P(\xi, \eta))^n\, d\xi\, d\eta\right)\right\}.$$

In order to obtain the numerical value of the probability of X having any specified value X', it remains only to specify the functions $p(n)$ and $P(\xi, \eta)$ and to use the familiar formula

$$(15) \qquad P\{X = X'\} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \phi(t) e^{-itX'}\, dt.$$

3. **Particular classes of the limiting distribution of** X. Until we have some experimental evidence as to what might be the nature of the two functions $p(n)$ and $f(x - \xi, y - \eta)$ or $P(\xi, \eta)$, we may try a few guesses. If the results

obtained in this way agree with empirical distributions, we shall have some reason to think that the guesses are not altogether wrong.

In certain cases all the larvae considered are at the moment of observation approximately of the same age. Alternatively, we may count only larvae which are at the same stage of development. With such counts it is not unreasonable to try for $p(n)$ either the binomial or the Poisson formula. Either of them will lead to easy calculations of (14). Writing

$$p(n) = e^{-\lambda} \frac{\lambda^n}{n!} \tag{16}$$

with λ representing the average number of survivors at the moment of observation per unit mass of eggs, we shall get for $\phi(t)$ the following expression

$$\phi(t) = \exp\left\{- Am\left(1 - \frac{1}{A} \int\int_A e^{\lambda P(\xi, \eta)(e^{it}-1)} \, d\xi \, d\eta\right)\right\}. \tag{17}$$

Substituting here for $P(\xi, \eta)$ any suitable function we shall obtain a corresponding particular form of the characteristic function $\phi(t)$, so that (17) determines a whole family of distributions. Substituting in (14) instead of (16), say the binomial formula, we shall obtain another family of contagious distributions.

Strictly speaking, in order to obtain some particular distribution from the formula (17), we have to specify the function $f(x - \xi, y - \eta)$, then to calculate $P(\xi, \eta)$ and substitute it in (17). Since however we have no knowledge of the properties of $f(x - \xi, y - \eta)$ and have to select it only on intuitive grounds, we may as well select the function $P(\xi, \eta)$. It may be selected either by itself directly, in which case there will be no difficulty in substituting it in (17), or by some indirect method. In the other case we may find it more convenient to use another form of (17) which is obtained by expanding the exponential under the sign of the integral in (17) and by integrating term by term, which is obviously permissible. In this way we get

$$\log \phi(t) = Am \sum_{n=1}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n. \tag{18}$$

Where P_n stands for the expression

$$P_n = \frac{1}{A} \int\int_A P^n(\xi, \eta) \, d\xi \, d\eta \tag{19}$$

and has the form of a moment of nth order of a certain probability law which it is easy to determine.

We may consider for a moment the value of $P(\xi, \eta)$ as a random variable Z. Its values cannot exceed the limits, zero and unity. Let z be any number between zero and unity and denote by $AF(z)$ the measure of the set of points belonging to A where $P(\xi, \eta) \leq z$. Then the function $F(z)$ will possess all the properties of the integral probability law of a variable Z which we may identify

with $P(\xi, \eta)$ and the integrals P_n will be simply the moments of Z namely, $P_n = \int_0^1 z^n dF$, where, of course, the integral would be considered in the sense of Stieltjes. It is interesting to notice that P_1 is always equal to A^{-1}. To see this consider the integral

$$AP_1 = \int\int_A P(\xi, \eta)\, d\xi\, d\eta \tag{20}$$

and substitute in it the expression of $P(\xi, \eta)$ in terms of the function $f(x - \xi, y - \eta)$. We get

$$AP_1 = \int\int_A d\xi\, d\eta \int\int_P f(x - \xi, y - \eta)\, dx\, dy \tag{21}$$

$$= \int\int\int\int_W f(x - \xi, y - \eta)\, dx\, dy\, d\xi\, d\eta. \tag{22}$$

Where the four-dimensional region of integration W is defined as follows. (i) The variables x and y vary so that the point having them for its coördinates may have any position within, but cannot be outside, of the experimental plot P. (ii) When x and y are fixed in the above way, say $x = x'$ and $y = y'$, then ξ and η may assume all those values for which the function $f(x' - \xi, y' - \eta)$ is positive. Let us denote this system of values of ξ and η by $B(x', y')$. Then we can calculate AP_1 as follows

$$AP_1 = \int\int_P dx\, dy \int\int_{B(x,y)} f(x - \xi, y - \eta)\, d\xi\, d\eta. \tag{23}$$

Now it is easy to see that the second integral in (23) is always equal to unity, whatever be x and y satisfying (i). To see this we have to recall the fundamental property of the function $f(x - \xi, y - \eta)$, due to the fact that it is the elementary probability law of x and y, namely that if ξ and η are fixed in one way or another, and it is integrated with respect to the other pair of variables, over all their values for which it is positive, the result will be equal to unity. In particular we shall have

$$\int\int_{f>0} f(u, v)\, du\, dv = 1. \tag{24}$$

Consider now the second integral in (23) and make the substitution

$$\xi = x - u, \qquad \eta = y - v \tag{25}$$

so that, instead of ξ and η we shall now integrate for u and v. It will be seen that the result of this substitution is exactly the integral (24), equal to unity. Since it was assumed that the area of P is equal to unity, it follows that $AP_1 = 1$. This equality is thus the necessary condition that the function $P(\xi, \eta)$ must satisfy. Besides, being a probability, it cannot be negative and cannot exceed unity. Whether any function having these properties may play the rôle of

$P(\xi, \eta)$ must be left for further inquiry. Assuming temporarily that this is so we can tentatively specify the probability laws belonging to the class determined by (18) by substituting in (18) instead of the P_n's the corresponding moments M_n of any distribution function $F(z)$ with its range between zero and unity, remembering only the interpretation of its first moment that we have found above, namely $M_1 = P_1 = A^{-1}$.

4. Certain general properties of the distributions deduced. Using the above result, we may substitute it in the formula (18) and get

$$\log \phi(t) = m\lambda(e^{it} - 1) + Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n. \tag{26}$$

Owing to the fact that the first term in the right hand side, $m\lambda(e^{it} - 1)$, represents the logarithm of the characteristic function of the Poisson Law,

$$p(x) = e^{-m\lambda} \frac{(m\lambda)^x}{x!} \tag{27}$$

for $x = 0, 1, 2, \cdots$ the formula (26) is especially interesting. Comparing the formulae

$$\begin{cases} P_1 = \displaystyle\int_0^1 z\, dF = A^{-1} \\ P_n = \displaystyle\int_0^1 z^n\, dF \end{cases} \tag{28}$$

we see that $0 < P_n \leq A^{-1}$ so that $AP_n \leq 1$. This circumstance assures the absolute and uniform convergence of (26). Frequently the higher moments P_n will be much smaller than the first, P_1, and if this tends to zero, all the products AP_n for $n \geq 2$ will do so too. In those cases $\log \phi(t)$ will tend to $m\lambda(e^{it} - 1)$ uniformly for all values of t. To see this take an arbitrary $\epsilon > 0$ and select N so large that

$$m \sum_{n=N+1}^{\infty} \frac{(2\lambda)^n}{n!} < \frac{\epsilon}{2}. \tag{29}$$

Next let A_0 be large enough for

$$AP_n < \frac{\epsilon}{2m} e^{-2\lambda} \tag{30}$$

for all $n = 2, 3, \cdots N$ and for any $A \geq A_0$. For such values of A we shall have

$$\begin{aligned} &\left| Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)}{n!} P_n \right| \\ &\qquad \leq Am \left(\sum_{n=2}^{N} \frac{\lambda^n \,|\, e^{it} - 1 \,|^n}{n!} P_n + \sum_{n=N+1}^{\infty} \frac{\lambda^n \,|\, e^{it} - 1 \,|^n}{n!} P_n \right) < \epsilon \end{aligned} \tag{31}$$

independently of what is the value of t. This result may be formulated as

PROPOSITION I. *If the parameters m and λ remain constant but the probability law $F(z)$ is changed so that all the products AP_n tend to zero for $n = 2, 3, \cdots$, then $\phi(t)$ tends to $m\lambda(e^{it} - 1)$ uniformly for all values of t and, consequently, the corresponding probability law of X tends to that of Poisson, given by* (27).

The above proposition may be considered as an explanation of the circumstance that occasionally the distribution of larvae may be very close to that of Poisson. This may happen for instance when the larvae that we count are sufficiently old and have had a sufficient time to travel very far from the spot where they were hatched. In such cases A will be large and, if the function $f(x - \xi, y - \eta)$ has some appropriate properties, all the products AP_n may be very small. But it is interesting to notice that there is a possibility of A increasing without the products AP_n tending to zero. Such will be for instance the case if $P(\xi, \eta)$ could have within A only two values $B_1(A)$ and $B_2(A)$ changing with A, one close to unity and the other close to zero. If Ap and Aq are the areas of the parts of A where $P(\xi, \eta)$ has those two different values, then we shall have

$$\begin{cases} P_1 = pB_1(A) + qB_2(A) = A^{-1} \\ P_n = pB_1^n(A) + qB_2^n(A) \end{cases} \tag{32}$$

and

$$AP_n = \frac{pB_1^n(A) + qB_2^n(A)}{pB_1(A) + qB_2(A)} \tag{33}$$

may tend to unity as A is increased. In such cases the probability law of X will not tend to (27). While calling attention to this possibility, it should be emphasized that it is not likely to occur in practice. In the cases of discontinuous $F(z)$ considered below $P\{X\}$ does tend to (17). The same is true also in such cases where it is assumed that

$$\begin{aligned} \frac{dF}{dz} &= a + bz \geq 0 \qquad \text{for} \quad 0 < 2 < c \leq 1 \\ &= 0 \qquad\qquad\quad \text{elsewhere} \end{aligned} \tag{34}$$

etc.

Before proceeding to specialize the expression (26) of the logarithm of the characteristic function, we shall show the connection existing between the P_n's and the semi-invariants of X. To calculate the latter it is sufficient to differentiate (26) with respect to t, to put $t = 0$, and to divide the result by the appropriate power of i. Denoting by γ_k the kth semi-invariant, by μ_1' the first moment about zero, and by μ_k the kth central moment of X we easily get

$$\begin{cases} \mu_1' = \gamma_1 = m\lambda \\ \mu_2 = \gamma_2 = m\lambda(1 + A\lambda P_2) \\ \mu_3 = \gamma_3 = m\lambda(1 + 3A\lambda P_2 + A\lambda^2 P_3) \\ \mu_4 - 3\mu_2^2 = \gamma_4 = m\lambda(1 + 7A\lambda P_2 + 6A\lambda^2 P_3 + A\lambda^3 P_4) \end{cases} \tag{35}$$

etc.

It will be seen that, in general, the kth semi-invariant depends on P_2, P_3, $\cdots P_k$ only. Another property of the new distributions that we shall mention is that they are "stable".

PROPOSITION II. *If $X_1, X_2, \cdots X_s$ are s independent random variables all following the same distribution with the logarithm of the characteristic function given by* (26), *then the sum $Y = \sum_{i=1}^{s} X_i$ will follow the same probability law with the exception that instead of the parameter m it will depend on the product sm.*

In order to establish this proposition it is sufficient to notice that the logarithm of the characteristic function of the variable Y is equal to the expression (26) multiplied by s.

Lastly, it may be noticed that the family of distributions determined by (26) is different from the comparable distributions deduced by Pólya ([3], p. 153, formulae (40) and (41)). In fact the logarithms of the characteristic functions of the latter could be written as follows:

$$-a \log (1 - b(e^{it} - 1)) = ab(e^{it} - 1) + a \sum_{n=2}^{\infty} \frac{b^n(e^{it} - 1)^n}{n} \tag{36}$$

and

$$\frac{c(e^{it} - 1)}{1 - de^{it}} = \frac{c(e^{it} - 1)}{1 - d} + \frac{c}{1 - d} \sum_{n=2}^{\infty} \left(\frac{d}{1 - d}\right)^{n-1} (e^{it} - 1)^n \tag{37}$$

respectively and, even if the formal expansions in powers of $(e^{it} - 1)$ converge, the identification of those expansions with (26) would require that P_n possess values exceeding unity, which is inconsistent with their essential property of being successive moments of a positive variable $0 \leq Z \leq 1$. Of course, the convergence of (36) and (37) would impose special restrictions on the constants that those formulae involve.

5. **Contagious distribution of type A depending on two parameters**. The simplest assumption that we can make concerning the function $P(\xi, \eta)$ is that it possesses some constant positive value within A and is zero elsewhere. Owing to (20) this constant value must be equal to A^{-1}. Substituting this in (17) we immediately obtain, say

$$\phi_1(t) = \exp\left\{-Am\left[1 - \exp\left(\frac{\lambda}{A} (e^{it} - 1)\right)\right]\right\}. \tag{38}$$

We could use the above formula directly to obtain the corresponding probability law. But before doing so, it may be useful to illustrate the machinery of the alternative method of obtaining the characteristic function of X and to calculate the same formula using (26).

If $P(\xi, \eta)$ is equal to A^{-1} everywhere in A, this means that the function $F(z)$ is a step function, which is equal to zero for any $z < A^{-1}$ and is equal to unity

elsewhere. Accordingly we shall have $M_n = A^{-n}$. Substituting this into (26) instead of P_n we easily get

$$\log \phi_1(t) = Am\left(e^{\frac{\lambda}{A}(e^{it}-1)} - 1\right) \tag{39}$$

which is equivalent with (38).

We shall now proceed to the calculation of the probabilities $P\{x = k\}$ as determined by either (38) or (39). For this purpose it will be useful to notice that the characteristic function (38) depends really on two parameters only, which we shall denote by m_1 and m_2,

$$m_1 = Am, \qquad m_2 = \lambda/A \tag{40}$$

In order to simplify the printing we shall further denote

$$z = m_1 e^{-m_2} \tag{41}$$

Expanding the two first exponentials of the three involved in (38), we may write

$$\phi_1(t) = e^{-m_1} \sum_{k=0}^{\infty} \frac{m_2^k}{k!} e^{ikt} \sum_{n=0}^{\infty} \frac{z^n}{n!} n^k. \tag{42}$$

This is the form of the characteristic function which is the most convenient when we have in mind applying the formula (15). In fact, it will be seen that we may multiply (42) by $e^{-iX't}$ and then integrate the series term by term. Further, it will be noticed that, on integrating between the limits $-\pi$ and $+\pi$, all the terms of the product will vanish except for the one which is independent of t. Consequently, the result of substituting (42) in the right hand side of (15) will be the coefficient of $e^{iX't}$ in the expansion (42), so that

$$P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \sum_{t=0}^{\infty} \frac{z^t}{t!} t^k. \tag{43}$$

As it is easy to verify, we have

$$P\{x = 0\} = e^{-m_1(1-e^{-m_2})} \tag{44}$$

and, for $k \geq 1$

$$P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \frac{d^k}{du^k} e^{m_1 e^{u-m_2}} \bigg|_{u=0}. \tag{45}$$

This formula gives an easy check of the identity $\sum_{n=0}^{\infty} P\{x = n\} = 1$. In fact, the left hand side can be looked upon as a product of e^{-m_1} by the Taylor's expansion of the function differentiated in (45) taken at the point $u = m_2$, which gives identically unity.

Successive differentiations give in turn

$$P\{X = 1\} = e^{-m_1(1-e^{-m_2})} \frac{m_2}{1!} m_1 e^{-m_2} \tag{46}$$

$$P\{X = 2\} = e^{-m_1(1-e^{-m_2})} \frac{m_2^2}{2!} (m_1^2 e^{-2m_2} + m_1 e^{-m_2}) \tag{47}$$

etc. Comparing the formulae (44), (46) and (47), the effect of the "contagiousness" of the distribution is easily seen. $P\{x = 2\}$ differs from what it would have been, if the distribution was that of Poisson, by the additional term $m_1 e^{-m_2}$ within the brackets.

Formulae (44), (46) and (47), and others which could be obtained by differentiating as indicated in (45), could be used for numerical calculations. However, these are greatly simplified by the use of the following elegant formula, deduced by Dr. Geoffrey Beall of the Dominion Entomological Experimental Station, Chatham, Ontario.

$$P\{X = n + 1\} = \frac{m_1 m_2 e^{-m_2}}{n + 1} \sum_{t=0}^{n} \frac{m_2^t}{t!} P\{X = n - t\}. \tag{48}$$

The correctness of this formula may be easily checked by calculating $P\{X = n - t\}$ from (43) and by substituting it in (48). Simple rearrangements will then give what could be obtained from (43) by putting $k = n + 1$.

Substituting $P_n = A^{-n}$ in formulae (35) and taking account of (40), we get

$$\mu_1' = \lambda m = m_1 m_2 \tag{49}$$

$$\mu_2 = \lambda m\left(1 + \frac{\lambda}{A}\right) = m_1 m_2(1 + m_2). \tag{50}$$

Solving these equations for m_1 and m_2 we obtain the formulae

$$m_2 = (\mu_2 - \mu_1')/\mu_1', \qquad m_1 = \mu_1'/m_2 \tag{51}$$

If the moments μ_1' and μ_2 are determined for an empirical distribution, these formulae may be used for estimating m_1 and m_2. In cases which were tried, this process did give frequently a satisfactory fit. Sometimes, however, when the tail of the original empirical distribution was very irregular, this distribution was better approximated by calculating the moments μ_1' and μ_2 not from itself but after a certain amount of smoothing of the tail. It follows that the method of fitting the new distribution to the empirical data requires some further study. At present it will suffice to mention that, whenever this distribution was tried on distributions of larvae which at the moment of counts were approximately of the same stage of development, the fit obtained was very satisfactory. It is hoped that a number of actual distributions fitted, together with the description of the method of counting, etc., will be soon published by Dr. Beall. As a matter of illustration one of his distributions is reproduced at the end of the present paper.

As for the distribution considered we have

$$\lim_{A\to\infty} AP_n = \lim A^{-n+1} = 0, \qquad n = 2, 3, \cdots \tag{52}$$

It follows from the above theory that, as $A \to \infty$, the probability law (48) tends to that of Poisson, namely

$$\lim_{A\to\infty} P\{X = n\} = e^{-m_1m_2}\frac{(m_1m_2)^n}{n!}. \tag{53}$$

For this reason the distribution (48) could be perhaps called the generalized probability law of Poisson, but it seems that the term "contagious distribution of type A with two parameters" will be more descriptive. Further on we shall see what is the justification of the description "of type A".

It was stated at the outset of the present paper that, when comparing the distributions of larvae in two series of plots subjected to two different treatments, there is sometimes doubt whether the means of those distributions are equal or not, while the difference in variability is more or less obvious. The formulae (49) and (50) give us the explanation of these facts. It is seen from the formula (49) that the mean of the distribution is equal to the product of the mean number of masses of eggs per unit of area and of the mean number of larvae per mass of eggs surviving at the moment of counts. If the two treatments compared are of about the same efficiency of killing the larvae, then the values of λ for each of them will be approximately equal and, consequently, we shall obtain about the same values for the two means. But while being of an equal efficiency as far as the killing is concerned, the two treatments may annoy the larvae in an unequal way. For example if the first treatment is dummy (no treatment) and the other is in general ineffective, it may still spoil the taste of the leaves that the larvae feed on. In such a case they may be compelled to travel a little more than they would otherwise, which will lead to an increase in A. Looking at the formula (50), it is easy to see that this would lead to a decrease in the value of μ_2. Alternatively the treatment may produce a temporary paralysis of the larvae which may reduce A and bring an increase of μ_2.

These remarks were applied to moments (49) and (50) of the particular distribution (45), but looking at the formulae (35), it is easily seen that they are true in the general case also.

6. Contagious distributions of type A depending on three parameters. As mentioned before, in order to determine some particular contagious distribution contained in the class depending on equation (18) it is sufficient to substitute in it instead of the P_n the moments of any distribution with its range confined to the interval from zero to unity, with the only restriction that the reciprocal of the first moment should be equal to A. Obviously this could be done in an infinity of ways, all of which will give more or less different results. We shall select the following one, representing a natural generalization of the procedure adopted above and leading to very simple formulae.

Formerly we have assumed that $P(\xi, \eta)$ possesses a constant value A^{-1} within the whole area A. At present we may assume that within this area it may possess one of two (three, four, etc.) values, say B_1 and B_2. Considering again $P(\xi, \eta)$ as a random variable Z, this will be equivalent to an assumption that Z may possess only one of the values B_1 and B_2 both positive and not exceeding unity. Again the probabilities of $Z = B_i$ are at our disposal. We shall take that these probabilities are equal, i.e. equal to $\frac{1}{2}$.

Comparing these assumptions with what may be the actual situation, one may be led to think that they are rather artificial. This however is not so. There is no doubt that the value of $P(\xi, \eta)$ does change within A, and it is also probable that the change is smooth. As we have no knowledge of the character of this function we first take its mean value within the area A and treat it as its first approximation. Next we divide the area A into two equal parts, say A_1 and A_2 and so that the greatest value of $P(\xi, \eta)$ in A_1 does not exceed any of the values in A_2. Then taking the average of $P\{\xi, \eta\}$ within A_1 and a similar average within A_2 and denoting them by B_1 and B_2 respectively, we do obtain a better approximation to the actual values of $P(\xi, \eta)$ assuming that it is equal to B_i everywhere in A_i. That is, in fact, the real meaning of the hypothesis formulated above and that we are going to accept in the following.

Denoting again by M_n the moments of Z we shall have

$$M_1 = \tfrac{1}{2}(B_1 + B_2) = A^{-1} \tag{54}$$

and generally

$$M_n = \tfrac{1}{2}(B_1^n + B_2^n). \tag{55}$$

Substituting (55) in (26) we get, say

$$\phi_2(t) = \frac{Am}{2}\left(e^{\lambda B_1(e^{it}-1)} + e^{\lambda B_2(e^{it}-1)} - 2\right). \tag{56}$$

We notice that this expression depends on three parameters, say

$$m_1 = Am, \qquad m_2 = \lambda B_1, \qquad m_3 = \lambda B_2. \tag{57}$$

In order to get the formulae for the probabilities of X having any specified values we could again apply the method used above when treating the more simple case. It may be useful however to illustrate a shorter way which easily leads to a generalization of Dr. Beall's recurrence formula. As we have noticed before, the probability $P\{X = k\}$ is equal to the coefficient of e^{ikt} in the expansion of the characteristic function in powers of e^{it}. Substituting for simplicity $z = e^{it}$, so that $t = -i \log z$, we may say that, if $\phi(t)$ is the characteristic function of a variable X_1 which is able to possess only integer values, then $P\{X = k\}$ is equal to the coefficient of z^k in the expansion of say $\psi(z) = \phi(-i \log z)$. Applying this rule to (56) we can write the following expression for the generating function $\psi(z)$,

$$\psi(z) = e^{-m_1} e^{\frac{1}{2}m_1\{e^{m_2(z-1)} + e^{m_3(z-1)}\}} = \sum_{k=0}^{\infty} z^k P\{X = k\}. \tag{58}$$

In other words

$$(59) \qquad P\{X = 0\} = \psi_{(0)} = e^{-m_1} e^{\frac{m_1}{2}\{e^{-m_2}+e^{-m_3}\}}$$

$$(60) \qquad P\{X = k\} = \frac{1}{k!}\frac{d^k\psi}{dz^k}\bigg|_{z=0}, \qquad k = 1, 2, \cdots$$

But

$$(61) \qquad \frac{d\psi}{dz} = \frac{m_1}{2}\psi(z)\,\{m_2 e^{m_2(z-1)} + m_3 e^{m_3(z-1)}\}$$

$$= \frac{m_1}{2}\psi(z)\chi(z) \qquad \text{(say)}$$

and it is easy to see that generally

$$(62) \qquad \frac{d^k\chi}{dz^k} = m_2^{k+1} e^{m_2(z-1)} + m_3^{k+1} e^{m_3(z-1)}.$$

As the kth derivative of $\psi(z)$ in (60) may be calculated by applying the familiar formula for the $(k - 1)$st derivative of the product $\psi(z)\chi(z)$ in (61), we obtain

$$(63) \qquad \frac{d^{n+1}\psi}{dz^{n+1}}\bigg|_{z=0} = \frac{m_1}{2}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\left(\frac{d^k\chi}{dz^k}\frac{d^{n-k}\psi}{dz^{n-k}}\right)\bigg|_{z=0}.$$

Using the formulae (60) and (62) we immediately obtain

$$(64) \qquad P\{X = n + 1\} = \frac{m_1}{2(n+1)}\sum_{k=0}^{n}\frac{m_2^{k+1}e^{-m_2} + m_3^{k+1}e^{-m_3}}{k!}P\{X = n - k\}.$$

As whenever $B_1 = B_2$ and consequently $m_2 = m_3$, the distribution considered now becomes identical with that considered formerly, depending on two parameters only, it is seen that the formula (64) represents a direct generalization of the formula (48). For purposes of successive calculation of the probabilities it will be probably more convenient to write (64) in the following form

$$(65) \qquad \begin{aligned} P\{X = n + 1\} = {} & \frac{m_1 m_2 e^{-m_2}}{2(n+1)}\sum_{k=0}^{n}\frac{m_2^k}{k!}P\{X = n - k\} \\ & + \frac{m_1 m_3 e^{-m_3}}{2(n+1)}\sum_{k=0}^{n}\frac{m_3^k}{k!}P\{X = n - k\}. \end{aligned}$$

This device of finding a recurrence formula for the probabilities will always succeed whenever there are no difficulties in finding the value of the nth derivative of the function χ.

It may be easily shown that if m and λ remain fixed but A tends to infinity, then the distribution (60) tends to the Poisson Law of frequency. Owing to

the general result stated in Proposition I, in order to show this it is only sufficient to prove that for $n \geq 2$

$$\lim_{A\to\infty} AM_n = \lim_{A\to\infty} \frac{B_1^n + B_2^n}{B_1 + B_2} = 0. \tag{66}$$

As both B_1 and B_2 must be included between zero and unity and their sum is equal to $2A^{-1}$, it follows that

$$0 < B_1 \leq A^{-1} \leq B_2 < 2A^{-1}. \tag{67}$$

Therefore

$$0 < AM_n < \frac{1 + 2^n}{2} A^{-n+1} \tag{68}$$

and (66) becomes obvious.

Substituting the values of M_2 and M_3 instead of P_2 and P_3 in the general expressions (35) of the moments, and taking into account the formulae (57), we obtain

$$\begin{cases} \mu_1' = \frac{1}{2}m_1(m_2 + m_3) \\ \mu_2 = \frac{1}{2}m_1(m_2 + m_3 + m_2^2 + m_3^2) \\ \mu_3 = \frac{1}{2}m_1(m_2 + m_3 + 3(m_2^2 + m_3^2) + m_2^3 + m_3^3). \end{cases} \tag{69}$$

If it is desired to fit the distribution to some empirical one using the method of moments, then these formulae could be solved with respect to m_1, m_2 and m_3. We may proceed as follows. Write

$$a = 2\mu_1^1, \qquad b = 2(\mu_2 - \mu_1'), \qquad c = 2(\mu_3 + 3\mu_2 + 2\mu_1'). \tag{70}$$

Then

$$m_1(m_2 + m_3) = a \tag{71}$$

$$m_1(m_2^2 + m_3^2) = b \tag{72}$$

$$m_1(m_2^3 + m_3^3) = c. \tag{73}$$

Multiplying the first of these equations by m_2 and subtracting the result from the second and repeating the same process with the second equation and the third, we get

$$\begin{aligned} m_1m_3(m_3 - m_2) &= b - am_2 \\ m_1m_3^2(m_3 - m_2) &= c - bm_2 \end{aligned} \tag{74}$$

and it follows

$$m_3 = \frac{c - bm_2}{b - am_2} \tag{75}$$

or

$$\frac{b}{a}(m_2 + m_3) - m_2 m_3 = \frac{c}{a}. \tag{76}$$

Again, dividing (73) by (71) we get

$$(m_2 + m_3)^2 - 3m_1 m_2 = \frac{c}{a}. \tag{77}$$

Multiplying (76) by 3 and subtracting from (77), we obtain

$$s^2 - 3bs/a - 2c/a = 0, \tag{78}$$

where $s = m_2 + m_3$. It follows that

$$s = \frac{3b}{2a} \pm \sqrt{\left(\frac{3b}{2a}\right)^2 - \frac{2c}{a}} \tag{79}$$

$$m_2 m_3 = p = \frac{bs - c}{a} \tag{80}$$

$$m_2 = \tfrac{1}{2}(s - \sqrt{s^2 - 4p}) \tag{81}$$

$$m_2 = \tfrac{1}{2}(s + \sqrt{s^2 - 4p}) \tag{82}$$

$$m_1 = a/s. \tag{83}$$

Following these steps we finally arrive to the values of all three parameters, given by the last three formulae.

If the values of the moments μ_1', μ_2 and μ_3 were known without error, the above formulae would give accurate values of m_1, m_2 and m_3. If, however, the moments are estimated from a sample, then the reader must be prepared that, even if the observed variable follows exactly the law, occasionally the sampling errors in the moments will make it impossible to carry out all the calculations indicated. Especially this may easily happen when the true values of m_2 and m_3 are equal or nearly equal, so that the empirical distribution is close to that given by the contagious distribution with only two parameters. As it is seen from (81) and (82), in such a case the true values of s and p must satisfy the relation

$$s^2 - 4p = 0. \tag{84}$$

However, the sampling errors in the moments will ascribe to the left hand side of (84) a value only approximately equal to zero, which may be either positive or negative. In the latter case we shall not be able to use (81) and (82) to estimate m_2 and m_3. As a matter of fact, the above circumstance actually arose in one case when it was tried to fit the three parameters distribution to a set of data which were excellently fitted by a simpler formula (45) involving only two parameters. As mentioned before, the problem of fitting the distributions which are deduced here requires further consideration.

Looking back on the method by which we have substituted a contagious distribution with three parameters m_1, m_2, m_3 for the simpler one with only two parameters, it is easily seen that it can be carried further leading to distributions with four, five, etc. parameters. In each case we would mentally divide the area A in a number of parts of equal size so that the values of $P(\xi, \eta)$ in the first never exceed those in the second, etc. Denoting the average values of $P(\xi, \eta)$ in those areas by $B_1, B_2, \cdots, B_r$, we shall obtain the moments

$$M_n = \frac{1}{r}\sum_{i=1}^{r} B_i^n, \tag{85}$$

substitute them in (26) and proceed more or less as we did above. All the distributions which may be obtained in this way possess certain common traits and I propose to call them "of type A". If the number of parameters in such a distribution is sufficiently high, it seems practically certain that the function $P(\xi, \eta)$ will be well approximated and we may hope to get an excellent fit. However, if a good fit may be attained only by introducing a great number of parameters, it usually means that the method of introducing those parameters is not very successful, and therefore it does not seem worth while to discuss in greater detail the distributions of type A with the number of parameters exceeding three. Instead we shall briefly indicate another class of distributions, built on another principle, which may be called of type B or C.

7. **Contagious distributions of types B and C.** As mentioned before, whenever the distributions of type A were tried on data, the character of which did not obviously contradict the basic assumptions of the theory (approximate equality of age of the larvae), the results were always satisfactory. However, our present experience is rather limited and it is well to anticipate the failures. We may expect that these will be caused by the over-simplified assumptions concerning the function $P(\xi, \eta)$.

In order to deal with such a case we may assume that for $0 < z < 1$ the derivative of $F(z)$ exists and is either a linear function of z or is equal to zero. Writing $p(z) = dF/dz$ we shall put

$$\begin{aligned} p_1(z) &= \tfrac{1}{2}A \quad \text{for } 0 < z < 2A^{-1}, \quad A \geq 2 \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{86}$$

Alternatively we may write, say

$$\begin{aligned} p_2(z) &= \frac{2A^2}{9}(3A^{-1} - z) \qquad \text{for } 0 < z < 3A^{-1} \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{87}$$

In the first case we shall obtain, say

$$M'_n = \frac{1}{n+1}\left(\frac{2}{A}\right)^n. \tag{88}$$

On the other hand, the moments of $p_2(z)$ will be given by

$$M''_n = \frac{2(3A^{-1})^n}{(n+1)(n+2)}. \tag{89}$$

Substituting these expressions in (26) we shall easily obtain the two new forms of the characteristic function of X, say

$$\log \phi_3(t) = -m_1 + m_1 \frac{e^{m_2(e^{it}-1)} - 1}{m_2(e^{it}-1)}, \tag{90}$$

with

$$m_1 = Am \quad \text{and} \quad m_2 = 2\lambda/A. \tag{91}$$

Accordingly, the generating function of the probabilities will be, say

$$\psi_3(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)}-1}{m_2(z-1)}} = \sum_{n=0}^{\infty} z^n P\{X = n\}. \tag{92}$$

The distribution determined by (92) may be called of type B.

Using the moments (89) and substituting them in the usual way in (26), we obtain, say

$$\log \phi_4(t) = -m_1 + 2m_1 \frac{e^{m_2(e^{it}-1)} - 1 - m_2(e^{it}-1)}{m_2^2(e^{it}-1)^2}, \tag{93}$$

with

$$m_1 = Am \quad \text{and} \quad m_2 = 3\lambda/A. \tag{94}$$

The probabilities of X having any specified value will be generated by the function, say

$$\psi_4(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)}-1-m_2(z-1)}{m_2^2(z-1)^2}} = \sum_{n=0}^{\infty} z^n P\{X = n\}. \tag{95}$$

The probability law determined by (95) may be called of type C. The comparative merits of all those distributions could be judged by comparing them with the results of observation.

8. **Illustrative Examples and Concluding Remarks.** Any series of positive terms adding up to unity may be considered as determining a probability law of a discontinuous variable such as the X considered above. When trying to obtain probability laws fitting the empirical distributions of some particular origin, the distributions of the numbers of larvae in experimental plots, or the like, we could really start by considering series of some positive terms each depending on one or more parameters, say

$$u_0(m_1, m_2),\ u_1(m_1, m_2),\ \ u_2(m_1, m_2),\ \cdots,\ u_n(m_1, m_2),\ \cdots \tag{96}$$

and having the property that, whatever the values of those parameters, $\sum_{n=0}^{\infty} u_n(m_1, m_2) = 1$. Studying a considerable number of empirical distributions,

we could apply the "method" of trial and error to guess the form of dependence of the $u_n(m_1, m_2)$ on the m's so that for a broad class of empirical distributions there would be a system of values of the m's, for which the series (96) would satisfactorily fit the data. If we succeed in this task we shall be entitled to a considerable satisfaction as the solution that we obtained would permit various further studies, e.g. the deduction of tests of significance applicable, or approximately applicable, in various cases, and so on.

Looking back at the history of statistics we shall find that the systems of frequency curves of Pearson, of Bruns-Charlier and others belong to the class of results just discussed. They are very important—and this especially applies to the Pearson curves—because of the empirical fact, that it is but rarely that we find in practice an empirical distribution, which could not be satisfactorily fitted by any of such curves. Consequently, wishing to deduce some test applicable in this or that case, we may usefully assume that the basic distribution is one of the Pearson system and, owing to the frequently continuous character of the connection between the conditions and the final results, our final formula will be approximately valid when applied to the data under consideration.

This point of view is not unfamiliar in pure mathematics. For example, we know that a broad class of functions may be approximated with any prescribed accuracy by means of polynomials. Wishing to prove a theorem applicable to this class of functions, we sometimes start by proving it for polynomials and then conclude that it is also true for the whole class. Here the rôle of polynomials is perfectly analogous to that of Pearson curves and could be described as that of good interpolation formulae.

But the problem of deducing theoretical distributions could be also considered from a slightly different point of view. Here again we require that the theoretical distribution fits satisfactorily the empirical data. But we may legitimately require something else: an "explanation" of the machinery producing the empirical distributions of a given kind. I have enclosed the word "explanation" in quotation marks so as not to suggest that I am attaching to it too much importance. Mathematics is always dealing with the conceptual sphere which is quite distinct from the perceptual and, at most, admits the possibility of establishing some correspondence. Therefore, however hard we try, we can never produce anything like a real mathematical explanation of any phenomena but instead only some "interpolation formula", some system of conceptions and hypotheses, the consequences of which are approximately similar to the observable facts. But this similarity may be differently placed. In the case of Pearson's curves it applies to the shape of these curves and to the shape of the empirical histograms. Otherwise it may apply to certain real features of the phenomena studied and to some mathematically described model of the same phenomena. And if the theoretical distributions deduced from the mathematical model do agree with those that we observe, and if that agreement is more or less permanent, we say that the mathematical model has "explained" the origin of the distributions.

If the problem of deducing interpolation formulae, sufficiently flexible to represent adequately a class of distributions, is of considerable interest, then that of producing similar formulae but involving an "explanation" of the phenomena studied, seems to be still more interesting. Of course, for it to be considered as successfully solved, the theoretical distributions deduced must fit the empirical ones, of a clearly specified kind, "practically always". At the present time we may quote a number of instances where it was possible to establish a mathematical probabilistic model of some class of phenomena determining probability laws which fit the empirical distributions with a remarkable accuracy. Perhaps the most important class of these phenomena is provided by the Mendelian theory; a number of other examples, although of a lesser importance but still interesting, have been mentioned elsewhere [2]. In all of them success-

TABLE I

Distribution of European corn borers in 120 groups of 8 hills each, (data provided by Dr. Beall), fitted by Poisson Law and by type A Law with two parameters

No. of borers	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	5.0	24	22.6
1	16.0	16	16.7
2	25.3	16	18.3
3	26.7	18	16.4
4	21.1	15	13.4
5	13.4	9	10.3
6	7.1	6	7.5
7	3.2	5	5.2
8	1.3	3	3.5
9	.4	4	2.3
10	.1	3	1.5
11		0	
12		1	
Beyond		—	2.3
m_1	—	—	2.178
m_2	—	—	1.454
P_{χ^2}	.000,000		> .95

TABLE II

Distribution of yeast cells in 400 squares of haemacytometer observed by "Student" (1907), fitted by Poisson Law and by type A Law with two parameters

No. of cells	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	202	213	214.8
1	138	128	121.3
2	47	37	45.7
3	11	18	13.7
4		3	3.6
5		1	.8
Beyond	2	—	.1
m_1	—	—	3.605
m_2	—	—	.189
P_{χ^2}	> .02		> .1

ful checks and rechecks increase our confidence that the conclusions based on the mathematical model determining the theoretical distributions will satisfactorily apply to observational data and also that our interpretations of various constants is more or less correct.

Now, what is the situation with the contagious distributions deduced above? They do represent an attempt to give good interpolation formulae involving an "explanation" of the observable phenomena, and all the constants introduced have meanings which are easy to interpret. Owing to the fact that in the process of the larvae surviving and spreading over the field there are certain unknown features, the final general formula that we have deduced involves two arbitrary functions $p(n)$ and $P(\xi, \eta)$. By substituting for them any appropriate functions that the intuition may suggest, we can obtain a number of distributions, each of which may or may not provide a satisfactory interpolation formula. Whether they do or not, must be empirically tested.

Up to the present time the contagious distributions of type A were tried on 12 distributions of larvae and on three distributions of yeast cells in squares of the haemacytometer, which did not quite agree with the Poisson Laws. The results of these trials were always the same: The type A distribution with two parameters provided an excellent fit, which was never worse than that of the more elaborate distribution with three parameters. This circumstance seems encouraging, but future experience may be less satisfactory and it would be very desirable to have some more empirical distributions and checks.

The following table gives two empirical distributions fitted with Poisson Law and with its generalization, as provided by the type A distribution with two parameters.

REFERENCES

[1] P. L. Hsu, "Contribution to the Theory of 'Student's' t-Test as Applied to the Problem of Two Samples." *Statistical Research Memoirs*, Vol. II, (1938) pp. 1–24.

[2] J. Neyman: "Lectures and Conferences on Mathematical Statistics," published by the Graduate School of the United States Department of Agriculture, Washington, D. C., 1938.

[3] G. Pólya: "Sur quelques points de la théorie des probabilités." *Annales de l'Institut Henri Poincaré*, Vol. 1, (1931) pp. 117–162.

[4] "Student": "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. 5, (1907), pp. 351–364.

University of California,
Berkeley, California.

ON CONFIDENCE LIMITS AND SUFFICIENCY, WITH PARTICULAR REFERENCE TO PARAMETERS OF LOCATION

By B. L. Welch

1. **Introduction.** The solution of the problem of estimating an interval in which a population parameter should lie, by means of what is now often termed the fiducial type of argument, dates back to the early writers on the theory of errors. However, owing to their lack of "Student's" z distribution, their statements were usually only of an approximate character, and, furthermore, the logical distinction between the fiducial method and the method of inverse probability was never clearly drawn, before R. A. Fisher discussed the subject. It is of interest to note how far "Student" himself went in this matter. In describing the tables which he gave in his original paper he says:[1]

"The tables give the probability that the value of the mean, measured from the mean of the population, in terms of the standard deviation of the sample, will lie between $-\infty$ and z. Thus, to take the tables for samples of six, *the probability of the mean of the population lying between* $-\infty$ *and once the standard deviation of the sample is* 0.9622 or the odds are about 24 to 1 that the mean of the population lies between these limits. The probability is therefore 0.0378 that it is greater than once the standard deviation, and 0.0756 that it lies outside $\pm$ 1.0 times the standard deviation."

It should be noted that "Student's" z is $(\bar{x} - \theta)/s$ where θ is the true population mean. His tables tell us that for $n = 6$, $P(z < 1)$[2] is equal to 0.9622. Owing to the symmetry of the z distribution this is equivalent to saying that $P(z > -1)$ is 0.9622, i.e.

$$P\left\{\frac{\bar{x} - \theta}{s} > -1\right\} = 0.9622.$$

This may be transposed to read

$$P\{\theta < \bar{x} + s\} = 0.9622 \tag{1}$$

which is the statement I have italicized in the above extract, it being there understood that the mean of the population is being measured from the mean of the sample. "Student" therefore makes here what is now called a fiducial statement. In the next sentence he, in effect, attaches a probability to an interval estimate for the population mean. In doing this "Student" was not conscious of introducing any new principle, nor does he apply the method consistently

[1] "Student" (1908). "The Probable Error of a Mean." *Biometrika* VI, p. 20.

[2] P is used to denote the probability of the truth of the relation in the bracket following.

to other problems of estimation. For instance, in discussing the estimation of the correlation coefficient ρ about the same time, he formulates the problem in terms of inverse probability, although he was fully aware of the difficulties involved in postulating an *a priori* distribution for ρ.

In discussing the problem of interval estimation more generally, I shall adopt some of the terminology used by J. Neyman.[3] The sample observations $x_1, x_2, \cdots x_n$ will be noted collectively by E (standing for the "event" point when the observations are represented as coördinates in a space of n dimensions). Then if θ is an unknown parameter, α a fixed probability, and $F(E, \theta, \alpha)$ a function such that

$$P\{F(E, \theta, \alpha) > 0\} = \alpha \tag{2}$$

we may obtain an interval estimate for θ as follows. Let $\delta(E, \alpha)$ denote the set of values of θ such that for any θ in the set we have $F(E, \theta, \alpha) > 0$. Then if we use the notation $\{\delta(E, \alpha) \subset \theta\}$ to indicate that the set $\delta(E, \alpha)$ contains or "covers" the true parameter θ we shall be able to rewrite (2)

$$P\{\delta(E, \alpha) \subset \theta\} = \alpha. \tag{3}$$

We can then adopt the following rule to obtain an interval estimate for θ: (a) calculate from the sample the set $\delta(E, \alpha)$, (b) make the statement that $\delta(E, \alpha)$ covers θ. In adopting this rule we shall be right in the proportion α of cases.

There are, in general, an infinite number of ways in which we can start with a statement of the type (2) to reach the statement of type (3). Neyman has discussed methods of making the best choice between such statements. His approach to this problem may be illustrated by the following example.

Suppose we have a random sample of n from a normal population with standard deviation σ and let

$$s^2 = \frac{\sum (x - \bar{x})^2}{(n - 1)},$$

and w = range = largest x − smallest x.
Then we can find a constant b_α such that

$$P\left\{\frac{s}{\sigma} > b_\alpha\right\} = \alpha \tag{4}$$

and, turning this round, we obtain

$$P\left\{\sigma < \frac{s}{b_\alpha}\right\} = \alpha. \tag{5}$$

This means that, if we choose $\alpha = .99$ (say), then we can say that σ is less than $s/b_{.99}$ and in 99% of cases we shall be correct in this statement.

---

[3] J. Neyman (1937). "Outline of a theory of statistical estimation based on the classical theory of probability." *Phil. Trans. Roy. Soc.* A 236, pp. 333–380.

Now similarly we can find c_α such that

$$P\left(\frac{w}{\sigma} > c_\alpha\right) = \alpha \tag{6}$$

and reversing this

$$P\left(\sigma < \frac{w}{c_\alpha}\right) = \alpha. \tag{7}$$

This statement is not inconsistent with (5). It means that, if we choose to base our rule of estimation *always* on the range, then in 99% of cases we shall be correct in saying that $\sigma < w/c_{.99}$. On the other hand, (5) relates to the consequences of applying *always* a rule of estimation based on the standard deviation of the sample. Both (5) and (7) are in themselves true statements, but we must decide which of them is the better one to use. In certain circumstances speed of calculation may be the determining factor, in which case (7) may be preferable, but here we shall assume that the time spent on calculation is not important.

In making the statement that σ is less than some upper limit which is a function of the sample observations, we shall, in general, prefer that this upper limit be placed as low as possible consistent with the chosen confidence coefficient α. We find, however, that it is not possible to say that, whatever the sample obtained, s/b_α will be less than w/c_α or *vice versa*. We must, therefore, approach the problem from another angle. If σ' is a value greater than the true standard deviation σ we can theoretically evaluate the *probability* that $\sigma' < s/b_\alpha$, and similarly the probability that $\sigma' < w/c_\alpha$. We may now express our general desire to place the upper confidence limit for σ as low as possible in a more concrete form. We may ask that the probability that σ' is less than this limit should be as small as possible. We find in the present problem that, whatever $\sigma' > \sigma$, we should include σ' in the interval from 0 to s/b_α less frequently than we should in an interval based on any other statistic. This constitutes an argument for using s rather than any other statistic such as w.

In general, Neyman makes all problems of choosing between alternative procedures of interval estimation depend on the probability that the intervals include values of the parameter different from the true value, as well as on the probability of them containing the true value. This principle of choice does, I think, appear reasonable, although its application is not, of course, so straightforward when statistics with properties of sufficiency similar to those of s do not exist. It is then necessary to introduce other conditions into the formulation of the problem. I intend to discuss elsewhere ways in which this has been done.

To summarize, we may say: (a) we can make many true statements of the type (3); and (b) if we can agree on certain further properties which these statements should possess, we can choose which is the best statement of this type to adopt as our general rule for interval estimation. There are certain differences

between this approach and that of R. A. Fisher, whose attitude is expressed clearly in his contribution to the discussion following Neyman's paper[4] "On the two different aspects of the representative method." Fisher says there that: "In particular he would apply the fiducial method, or rather would claim unique validity for its results, only in those cases for which the problem of estimation proper had been completely solved, i.e. either when there existed a statistic of the kind called *sufficient*, which in itself contained the whole of the information supplied by the data, or when, though there was no sufficient statistic, yet the whole of the information could be utilized in the form of *ancillary* information." Thus it appears that when sufficient statistics do not exist, excepting in those further cases where Fisher claims that the problem of estimation has been completely solved, he would definitely discourage the use of the fiducial argument at all. Neyman, on the other hand, would allow the attempt to obtain interval estimates on the lines described above. Where sufficient statistics do exist, the two approaches do not lead to any final disagreement. Neyman, using results obtained in the Neyman-Pearson theory of testing hypotheses, is led to criteria depending in a particular way on the joint probability law of the sample, and these criteria are seen to involve the sample values only through statistics which have been defined as sufficient. One may regard this fact in two ways: (a) one may say that because a certain line of approach, which seems intuitionally sound, leads to the use of statistics which have been defined as sufficient, therefore this definition of sufficiency is a good one, or (b) one may say that the definition of sufficient statistics is fundamental, and that any method of approach which leads to their use has thereby obtained some extra support.

There remains the case alluded to above, where the joint probability law of the sample does not depend on the unknown parameter θ by way of one statistic only, but where nevertheless it has been said that the problem of estimation has been completely solved. This case will be discussed in the next section.

2. **Interval Estimates of Location.** R. A. Fisher has given, as a particular example, a case where the unknown parameter is one of location, so that we can write

$$p(x \mid \theta) = \phi(x - \theta).$$

Now if we have a sample of n from this distribution, the $(n - 1)$ differences between successive observations when arranged in order of magnitude will have a joint distribution independent of θ. Hence if we denote the sample by E, and the $(n - 1)$ differences jointly by C, we have

$$p(E \mid \theta) = p(T \mid C, \theta)p(C) \tag{8}$$

where T is some statistic, such as the mean or median, whose distribution does depend on θ and may hence be taken as an estimate of θ. We may therefore

[4] J. Neyman (1934). *J. R. Statist. Soc.* 97, p. 617.

read (8) as follows: the joint probability law of the sample is equal to the probability law of the estimate in samples of the same configuration, C, multiplied by the probability of the configuration, the latter not depending on the unknown θ. From this it has been deduced that all the information respecting θ provided by the sample is given by referring T to the distribution $p(T \mid C, \theta)$. Fisher,[5] for instance, says that "in interpreting our estimate (we) may take as its sampling distribution that appropriate to only those samples which have the actual configuration observed." Later in the same context he remarks that in general, when θ is a parameter of any type whatever, and not necessarily one of location or scaling, if something can be found "corresponding with the configuration of the sample in the simple case discussed above, . . . one of the primary problems of uncertain inference will have reached its complete solution. If not, there must remain some further puzzles to unravel."

It is clear, therefore, that more has been claimed for this method than that it is *practically* useful, or that it yields the best results possible in *large* samples, or that it yields results *highly approximating* the best possible in small samples. There is an emphasis here on completeness that leads one to suppose that all problems of estimation and testing hypotheses may be answered to the best advantage by considering only the distribution of an estimate in samples of the same configuration, the estimate thus attaining properties analogous to those of a sufficient statistic. That this supposition is not true may be seen by considering the following simple example. This example concerns the simplest situation that one deals with in the theory of testing statistical hypotheses. Its relevance to the problem of interval estimation will, however, not be difficult to see.

Suppose that we have a sample from a population involving only a parameter of location θ, and that we wish to test whether θ is equal to θ_0 (say), and that besides θ_0 there is only one value θ_1 (say) which it is possible for θ to take. Suppose we require to set up a statistical test which will reject the hypothesis $\theta = \theta_0$, in only a small proportion ϵ of cases, when it is true. Many such tests are possible, and it is natural to choose from them that test which will lead most frequently to the rejection of the hypothesis that $\theta = \theta_0$ when the single alternative $\theta = \theta_1$ is true. Neyman and Pearson[6] have shown that the best test from this viewpoint is provided by the criterion

$$J = \frac{p(E \mid \theta_1)}{p(E \mid \theta_0)}. \tag{9}$$

This criterion must be referred to its distribution in *all* samples when $\theta = \theta_0$. We must therefore choose a constant J_ϵ such that

$$P(J > J_\epsilon \mid \theta = \theta_0) = \epsilon \tag{10}$$

[5] Fisher, R. A. (1936). "Uncertain Inference." *Proc. Amer. Acad. Arts and Sciences*, **71**, No. 4, p. 257.

[6] J. Neyman and E. S. Pearson (1932). "On the problem of the most efficient tests of statistical hypotheses." *Phil. Trans. Roy. Soc.* A 231, p. 300.

and reject the hypothesis that $\theta = \theta_0$ when $J > J_\epsilon$. This is known to be the best test in these circumstances, and we may demand that any other procedure which claims to use the data exhaustively should be equivalent to it. Now if we decide to use only the distribution of the statistic T in samples of the same configuration, we are led to take as the most powerful test based on $T \mid C$ one which would reject the hypothesis that $\theta = \theta_0$ when the ratio of $p(T \mid C, \theta_1)$ to $p(T \mid C, \theta_0)$ exceeds a certain value. Now by (8) this ratio is exactly the criterion J of (9) above. There is, however, this difference, that J has now to be referred to its distribution in samples with the *same configuration* C as that observed. We shall therefore have to choose $J_\epsilon(C)$ such that

$$P(J > J_\epsilon(C) \mid C, \theta\} = \epsilon. \tag{11}$$

A test, then, which rejects the hypothesis that $\theta = \theta_0$ when $J > J_\epsilon(C)$ will be such that it is the most powerful possible with respect to the alternative $\theta = \theta_1$, based on samples with the same configuration. However, in actual sampling from a population, we derive samples with all configurations, and the real power of the test will therefore be measured by

$$P\{J > J_\epsilon(C) \mid \theta_1\} = \int P\{J > J_\epsilon(C) \mid C, \theta_1\} p(C)\, dC. \tag{12}$$

This quantity cannot be greater, and will in general be less, than the power[7] of the other test, viz. $P(J > J_\epsilon \mid \theta_1)$. (If $J_\epsilon(C)$ is the same for all C, and therefore equal to J_ϵ, the powers will be equal. This will be the case when there is a sufficient statistic for θ.) We must therefore conclude that, in relation to this simple problem at least, a method which takes account only of distributions in samples with the same configuration will not use the data to the best advantage.

Of course the type of problem to be solved is usually not so straightforward as the present one. There will usually be more than one value of θ alternative to θ_0, and no uniformly most powerful test will, in general, exist. It is legitimate, however, to consider the above example, because any procedure claiming properties of sufficiency should be able to deal with it in the best possible way.

An example may make the above points clearer, and will show their relevance to the problem of interval estimation. Consider a rectangular distribution with mean θ, and range from $(\theta - \frac{1}{2})$ to $(\theta + \frac{1}{2})$. Let x_1 and x_2 be a sample of 2 from this population, and suppose we require confidence limits for θ such that the chance of them enclosing θ is α.

If we represent x_1 and x_2 as coördinates of a point with respect to rectangular axes, the joint probability distribution is constant over a square centered at the point (θ, θ). This is shown by $ABCD$ in Fig. 1. We have

$$p(x_1, x_2)dx_1dx_2 = dx_1dx_2 \begin{cases} \theta - \frac{1}{2} < x_1 < \theta + \frac{1}{2} \\ \theta - \frac{1}{2} < x_2 < \theta + \frac{1}{2}. \end{cases} \tag{13}$$

[7] Power is used throughout in the Neyman-Pearson sense, i.e. to denote the chance of a test rejecting a hypothesis when a given alternative is actually true.

If we write $z_1 = \frac{1}{2}(x_1 + x_2)$; $z_2 = \frac{1}{2}(x_1 - x_2)$, z_2 will represent the configuration of the sample, and z_1 may be taken as the estimate, T, of θ in our discussion above. We can then show that

$$p(z_1, z_2)dz_1dz_2 = 2dz_1dz_2, \tag{14}$$

$$p(z_2)dz_2 = 2\{1 - 2 \mid z_2 \mid \}dz_2 \cdots -\tfrac{1}{2} < z_2 < \tfrac{1}{2}, \tag{15}$$

and

$$p(z_1 \mid z_2)\, dz_1 = \frac{dz_1}{1-2 \mid z_2 \mid} \cdots \theta - \tfrac{1}{2} + \mid z_2 \mid < z_1 < \theta + \tfrac{1}{2} - \mid z_2 \mid . \tag{16}$$

That these are the correct limits for z_1 and z_2 may be seen by reference to Fig. 1, noting that z_1 and z_2 are constant along lines parallel to the respective diagonals BD and AC of the square.

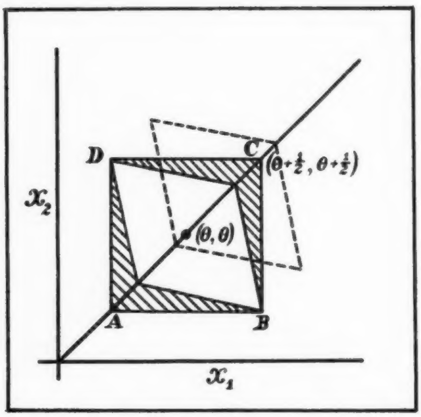


FIG. 1

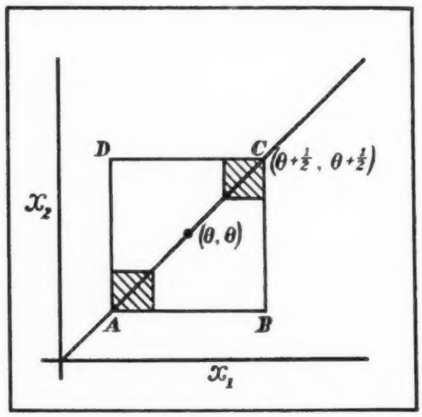


FIG. 2

First let us confine ourselves to samples with the same configuration z_2. Then, from (16), we can say that

$$P\{\theta - \alpha(\tfrac{1}{2} - \mid z_2 \mid) < z_1 < \theta + \alpha(\tfrac{1}{2} - \mid z_2 \mid)\} = \alpha. \tag{17}$$

This statement is true for given z_2, and will be *a fortiori* true when this restriction is removed. It is equivalent to saying that the chance of a point falling into the shaded area in Fig. 1 is $(1 - \alpha)$, where α denotes the proportion of the diagonal AC lying in the non-shaded area.[8] Confidence limits for θ are then obtained by transposing (17), giving

$$P\{z_1 - \alpha(\tfrac{1}{2} - \mid z_2 \mid) < \theta < z_1 + \alpha(\tfrac{1}{2} - \mid z_2 \mid)\} = \alpha. \tag{18}$$

[8] We are assuming that confidence limits are required such that the chance is $\left(\frac{1}{2} - \frac{\alpha}{2}\right)$ of θ being above the upper limit, and $\left(\frac{1}{2} - \frac{\alpha}{2}\right)$ of it being below the lower limit.

That this is not the best way of constructing confidence limits is seen as follows. Let us denote the lesser of x_1 and x_2 by x_L, and the greater by x_G. Then if we consider the possible values of x_L and x_G which will satisfy simultaneously the inequalities

$$(19)\qquad \begin{cases} \theta - \frac{1}{2} < x_L < \theta + \frac{1}{2} - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ \theta - \frac{1}{2} + \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} < x_G < \theta + \frac{1}{2} \end{cases}$$

we see that they lie in the non-shaded area of the square $ABCD$ in Fig. 2 where the sides of the shaded squares are $\sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}}$. The chance of the inequalities holding simultaneously is therefore α. Further we see that these inequalities can be transposed to read

$$(20)\qquad \begin{cases} x_G - \frac{1}{2} < \theta < x_L + \frac{1}{2} \quad \text{when } (x_G - x_L) > \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ x_L - \frac{1}{2} + \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} < \theta < x_G + \frac{1}{2} - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ \qquad\qquad\qquad\qquad \text{when } (x_G - x_L) < \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \end{cases}$$

and therefore we can take these to define our confidence limits for θ.

The intervals defined by the confidence limits in (18) and (20) are equivalent in the sense that each covers the true value of θ in a proportion α of cases. To decide which is the better rule of interval estimation we shall follow Neyman, and consider how often the intervals cover values other than the true θ. In particular let $(\theta + \Delta)$ be any other value, and consider the expressions P_1 and P_2 where

$$(21)\qquad P_1 = P\{z_1 - \alpha(\tfrac{1}{2} - |z_2|) < (\theta + \Delta) < z_1 + \alpha(\tfrac{1}{2} - |z_2|)$$

and P_2 is the probability that one or another of the following inequalities holds

$$(22)\qquad \begin{cases} x_G - \frac{1}{2} < (\theta + \Delta) < x_L + \frac{1}{2} \quad \text{when } (x_G - x_L) > \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ x_L - \frac{1}{2} + \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} < (\theta + \Delta) < x_G + \frac{1}{2} - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ \qquad\qquad\qquad\qquad \text{when } (x_G - x_L) < \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \end{cases}$$

Now (21) can be written

$$(23)\qquad P_1 = P\{(\theta + \Delta) - \alpha(\tfrac{1}{2} - |z_2|) < z_1 < (\theta + \Delta) + \alpha(\tfrac{1}{2} - |z_2|)\}.$$

Referring to Fig. 1 we see that we have to evaluate the chance of the sample falling into a lozenge-shaped area like the unshaded area in $ABCD$, but moved bodily along the diagonal AC to such a position as is indicated by the dotted lines. Difficulties are introduced by the discontinuities, but we can show that for Δ positive

$$(24)\quad \begin{cases} P_1 = \alpha \quad \text{when } \Delta = 0 \\ P_1 = \alpha - \dfrac{4\alpha\Delta^2}{1-\alpha^2} \cdots 0 \leq \Delta \leq \left(\dfrac{1}{2} - \dfrac{\alpha}{2}\right) \\ P_1 = \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) - 2\Delta + \dfrac{2\Delta^2}{1+\alpha} \cdots \left(\dfrac{1}{2} - \dfrac{\alpha}{2}\right) \leq \Delta \leq \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) \\ P_1 = 0 \qquad \Delta \geq \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) \end{cases}$$

with similar expressions for Δ negative. The graph of P_1 against Δ is shown in Fig. 3, α for convenience being taken $= 0.92$. From it we can read off the probability of the confidence interval covering $(\theta + \Delta)$, where θ is the true value of the parameter.

Similar calculations may be made for P_2. Without going into details, it is seen that

$$(25)\quad \begin{cases} P_2 = \alpha & \text{when } \Delta = 0 \\ P_2 = \alpha - 2\Delta\left(1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}}\right) & 0 \leq \Delta \leq \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ P_2 = \left(\dfrac{1}{2} + \dfrac{\alpha}{2}\right) - 2\Delta + \Delta^2 & \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \leq \Delta \leq 1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \\ P_2 = 0 & \Delta \geq 1 - \sqrt{\dfrac{1}{2} - \dfrac{\alpha}{2}} \end{cases}$$

P_2 is plotted against Δ in Fig. 3. It is seen that, whatever value of Δ we take, the chance of $(\theta + \Delta)$ being included in the confidence interval, is less for the second method of estimation than it is for the method based on the distribution of $z_1 \mid z_2$.[9] This circumstance would, I think, contradict the view that the latter method was deriving the utmost from the sample. Whether the method is still a good one, though not necessarily the best, is not a question at issue in the present paper. The curves in Fig. 3 are very close together, and we are led to expect this by the fact that (12) is the weighted mean of the powers within the separate configurations, the weights being the probabilities $p(C)$ of the configurations. I am only concerned to show that certain methods, for which

[9] It will be noted that, when inverted, the curves of Fig. (iii) represent the power functions of tests for which the regions of rejection are those in figures (i) and (ii) respectively, the test being whether the parameter has the specified value θ, and different alternative hypotheses being represented by $(\theta + \Delta)$.

properties analogous to those of sufficiency have been claimed, do not satisfy conditions which I think they should, if these claims are to be upheld.

3. **Fiducial Distributions.** In the first section of this paper I discussed certain points of difference between the approaches to the problem of interval estimation made by R. A. Fisher on the one hand and J. Neyman and E. S. Pearson on the other. The differences are not, perhaps, of the same magnitude as those between all these writers and the protagonists of inverse probability, and the results reached are so often the same that the reader may be excused for being somewhat impatient with what appear to be rather fine distinctions. However, as was seen in the last section, the approaches do not always yield exactly the

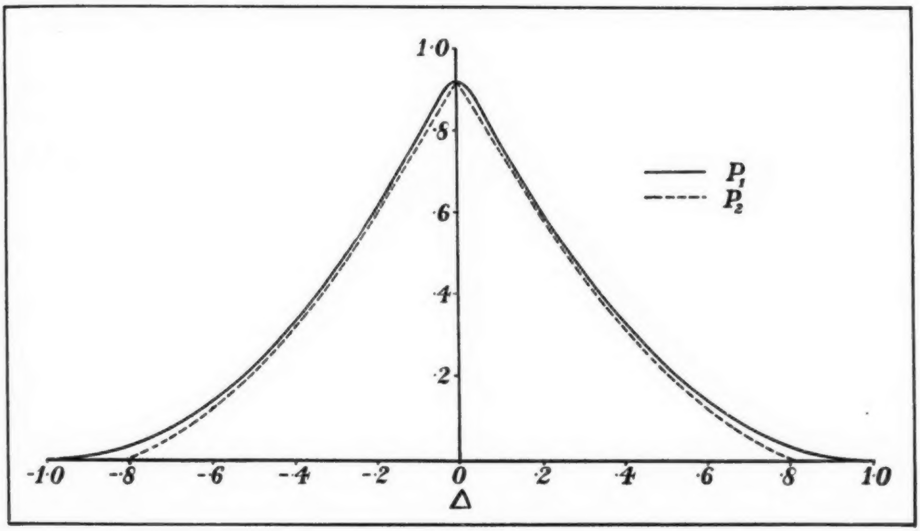


FIG. 3

same final results, and therefore I think it may be profitable to discuss them still further.

Closely connected with Fisher's desire to restrict the use of the fiducial method to situations where statistics exist which possess some property of sufficiency, is his introduction of the concept of a *fiducial distribution* for the unknown parameter. One can talk about *the* fiducial distribution for a parameter only if it is a unique distribution. Neyman, however, never makes use of fiducial distributions, and would, I think, claim that any valid results reached with the concept can equally well be reached without it. Where the results are the same there is room for two opinions on this matter. Some writers find it convenient to think in terms of fiducial distributions, and others prefer always to carry forward their reasoning as far as possible in terms of direct probability statements about the observational values, before transposing them to obtain confidence or fiducial limits for the parameters.

Greater objection can be made to the use of simultaneous fiducial distributions of several parameters. For instance, in the case of the normal distribution with parameters μ and σ, a simultaneous fiducial distribution has been defined in the following way.[10] Starting with the fact that the joint distribution of

$$\phi_1 = \frac{\sqrt{n}\,(\bar{x} - \mu)}{\sigma} \qquad \text{and } \phi_2 = \frac{(n-1)s^2}{\sigma^2}$$

is

$$df = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2})\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}\phi_1^2} e^{-\frac{1}{2}\phi_2} \phi_2^{\frac{1}{2}(n-3)}\, d\phi_1\, d\phi_2,$$

$\bar{x}$ and s are treated *formally* as fixed, and ϕ_1 and ϕ_2 are transformed to μ and σ, treated *formally* as variables. This gives

$$(26) \qquad df = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2})\Gamma\left(\frac{n-1}{2}\right)} \frac{\sqrt{n}}{\sigma} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} \cdot \frac{2}{\sigma} e^{-\frac{(n-1)s^2}{2\sigma^2}} \left\{\frac{(n-1)s^2}{\sigma^2}\right\}^{\frac{1}{2}(n-1)} d\mu\, d\sigma$$

This distribution would be useful if it were legitimate to integrate it out to obtain a fiducial distribution for any function $g(\mu, \sigma)$ say, of μ and σ. However, as for instance Bartlett has pointed out, this is not necessarily permissible. It seems to me therefore, that distributions defined as in (26) should be dispensed with entirely, for their very form encourages the belief that they can be integrated out at will. That this belief is still held is illustrated by a recent paper by Miss D. M. Starkey[11] concerned with the difference between the means of normal populations where the standard deviations are not assumed equal. This is the original problem to which Fisher[12] applied a method equivalent to integrating out the joint fiducial distribution of the two population means. Bartlett[13] raised an objection to this method of treatment, and I have also discussed the matter further.[14] Miss Starkey proceeds from the assumption that Fisher's method is sound.

The concept of the fiducial distribution has also been used in those problems of location and scaling, which have been treated by the procedure discussed above, of considering distributions in samples with the same configuration. Indeed it is one of the attractions of this procedure that we are led to distribu-

---

[10] R. A. Fisher, (1935). "The fiducial argument in statistical inference." *Ann. Eugen.* VI, p. 395.

[11] Daisy M. Starkey (1937). "A test of the significance of the difference between means of samples from two normal, populations without assuming equal variances." *Ann. Math. Stat.* Vol. IX. No. 3, pp. 201–213.

[12] R. A. Fisher (1935). *loc. cit.*

[13] M. S. Bartlett (1936). "The information available in small samples." *Proc. Camb. Phil. Soc.* 32, pp. 560–566.

[14] B. L. Welch (1937). "The significance of the difference between two means when the population variances are unequal." *Biometrika*, XXIX, p. 358.

tions with, so to speak, one degree of freedom, so that the fiducial method may be safely applied. However, although probability statements based on such a fiducial method are here quite valid, I do not think that such statements can claim a *unique* validity. As I have shown in the previous section, there is no necessity to confine oneself to sampling within a configuration in order to obtain interval estimates for parameters, and we may fare better by not so confining ourselves, even if we have to dispense with the fiducial distribution.

4. **Summary.** Certain points which arise in the problem of estimating an interval in which a population parameter should lie have been discussed. In the second section it has been shown that in estimating location parameters it is not sufficient to consider the distribution of estimates in samples of the same configuration, meaning by sufficient that the sample is thereby utilized in the best possible way.

University College,
London

THE REGRESSION SYSTEMS OF TWO SUMS HAVING RANDOM ELEMENTS IN COMMON

By J. F. Kenney

1. **Introduction.** The purpose of this note is to illustrate the power and elegance of the technique of characteristic functions[1] in solving a problem which has been discussed in the literature by Fischer[2] and others.

Let $x_1, x_2, \cdots, x_n$ be n variables independent of each other in the statistical sense, all subject to the same distribution function f, so that the function representing their joint distribution is

$$f(x_1)f(x_2)\cdots f(x_n). \tag{1}$$

Under these conditions a set of values $x_1, x_2, \cdots, x_n$ will be said to constitute a *sample of n* from a population with distribution function $f(x)$ and the function (1) will be said to represent the distribution of samples. It will be understood that $f(x)$ is defined and is non-negative for all real values of x and

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

If the actual occurrence of the variable is limited to a finite range, $f(x)$ is defined as identically zero outside that range.

The mathematical expectation of an arbitrary function $\psi(x)$, denoted by application of the operator E, is

$$E[\psi(x)] = \int_{-\infty}^{\infty} \psi(x)f(x)\,dx. \tag{2}$$

This integral will be convergent whenever $\psi(x)$ is absolutely integrable and bounded. In particular, if $\psi(x) = x$ we have the mean

$$a = \int_{-\infty}^{\infty} xf(x)\,dx$$

and it will be assumed that a exists.

Suppose a sample of n is taken from the population represented by $f(x)$ and the sum

$$y = x_1 + x_2 + \cdots + x_k + x_{k+1} + \cdots + x_n \tag{3}$$

---

[1] The writer takes pleasure in acknowledging his indebtedness to Professor A. T. Craig for suggesting this method.

[2] "On correlation surfaces of sums with a certain number of random elements in common," these Annals, vol. 4, no. 2, pp. 103–126.

is formed. From this sample $k < n$ values are chosen at random, and a sample of $m - k$ $(m \leq n)$ additional values, x_j', is taken from $f(x)$. The sum

$$z = x_1 + x_2 + \cdots + x_k + x_{k+1}' + \cdots + x_m' \tag{4}$$

is then formed. The problem is to determine the regression systems of z on y and y on z in the population resulting from repeated samples.

Before proceeding with the solution a brief discussion of characteristic functions will be given.

2. **Characteristic functions.** When $\psi(x) = e^{itx}$, where t is a real variable and $i = \sqrt{-1}$, (2) is called the characteristic function of x. Thus if we let $\varphi(t) = E(e^{itx})$ we have

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx.$$

From the conditions imposed on $f(x)$ it follows that the integral defining $\varphi(t)$ is convergent and $|\varphi(t)| \leq 1$. If the kth derivative of $\varphi(t)$ with respect to t exists we have

$$\left.\frac{d^k \varphi(t)}{dt^k}\right|_{t=0} = i^k \nu_k$$

where

$$\nu_k = \int_{-\infty}^{\infty} x^k f(x)\, dx.$$

Thus the characteristic function of x has the property that its kth derivative at the origin (divided by i^k) gives the kth moment of the distribution of x about the origin of x.

The notion of characteristic function extends readily to a distribution of several variables. In particular, let $F(y, z)$ be the joint distribution function of variables y and z subject to the condition

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(y, z)\, dy\, dz = 1.$$

Then the characteristic function of $F(y, z)$ is

$$\varphi(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it_1 y + it_2 z} F(y, z)\, dy\, dz \tag{5}$$

where y and z are defined in (3) and (4).

3. **Solution of the problem.** The distribution function associated with the population of samples is of the form given by (1). Consequently, the characteristic function of $F(y, z)$ can be written in the form

$$\int \cdots \int \prod_{j=1}^{k} e^{i(t_1+t_2)x_j} f(x_j)\, dx_j \prod_{j=k+1}^{n} e^{it_1 x_j} f(x_j)\, dx_j \prod_{j=k+1}^{m} e^{it_2 x_j'} f(x_j')\, dx_j'$$

the limits of integration being taken over all admissible values of the variables. The above expression reduces to

$$\varphi(t_1, t_2) = [\varphi(t_1 + t_2)]^k [\varphi(t_1)]^{n-k} [\varphi(t_2)]^{m-k}. \tag{6}$$

By the Fourier transform we have from (5),

$$F(y, z) = (1/2\pi)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-it_1y-it_2z} \varphi(t_1, t_2)\, dt_1\, dt_2.$$

Since a distribution is completely determined by its characteristic function, $F(y, z)$ can be exhibited if $f(x)$ is known. However, the solution of the problem does not depend upon exhibiting $F(y, z)$.

Let $g(y)$ and $h(z)$ be the marginal distributions of y and z, respectively. Then the mean value of z for a fixed y is

$$\bar{z}_y = \int \frac{zF(y, z)}{g(y)}\, dz, \tag{7}$$

and the mean value of y for a fixed z is

$$\bar{y}_z = \int \frac{yF(y, z)}{h(z)}\, dy \tag{8}$$

where here and subsequently the integration is taken over all admissible values of the variables.

Let us now take the partial derivative of $\varphi(t_1, t_2)$, as given in (5), with respect to t_2 and evaluate the result at $t_2 = 0$. We obtain

$$\frac{\partial}{\partial t_2} \varphi(t_1, t_2)\bigg|_{t_2=0} = \int \int ize^{it_1y} F(y, z)\, dy\, dz. \tag{9}$$

If we denote the left member of (9) by $G(t_1)$ and utilize (7) in the right member, (9) becomes

$$G(t_1) = \int g(y)\bar{z}_y ie^{it_1y}\, dy.$$

Application of the Fourier transform yields

$$ig(y)\bar{z}_y = \frac{1}{2\pi} \int e^{-it_1y} G(t_1)\, dt_1. \tag{10}$$

Now from (6),

$$G(t_1) = k\overline{\varphi(t_1)}^{n-1} \varphi'(t_1) + \overline{\varphi(t_1)}^n ia(m - k).$$

Therefore (10) may be written as follows,

$$ig(y)\bar{z}_y = \frac{k}{2\pi} \int e^{-it_1y}\overline{\varphi(t_1)}^{n-1} \varphi'(t_1)\, dt_1 + \frac{ia(m - k)}{2\pi} \int e^{-it_1y}\overline{\varphi(t_1)}^n\, dt_1. \tag{11}$$

To evaluate these integrals, consider

$$\overline{\varphi(t_1)}^n = \int e^{it_1y} g(y)\, dy. \tag{12}$$

Differentiating (12) with respect to t_1 we have

$$n\overline{\varphi(t_1)}^{n-1}\varphi'(t_1) = \int iye^{it_1y} g(y)\, dy. \tag{13}$$

Again using the Fourier transform, we obtain

$$iyg(y) = \frac{nk}{2k\pi} \int e^{-it_1y}\overline{\varphi(t_1)}^{n-1}\varphi'(t_1)\, dt_1$$

from (13) and

$$g(y) = \frac{1}{2\pi} \int e^{-it_1y}\overline{\varphi(t_1)}^n\, dt_1$$

from (12). Therefore (11) reduces to

$$iyg(y)\bar{z}_y = \frac{k}{n} iyg(y) + ia(m - k)g(y)$$

and we have at once the simple result

$$\bar{z}_y = ky/n + a(m - k). \tag{14}$$

In an analogous manner, it may be shown that

$$\bar{y}_z = kz/m + a(n - k). \tag{15}$$

Writing (14) and (15) in the forms

$$\begin{cases} \bar{z}_y - am = c_1(y - an) \\ \bar{y}_z - an = c_2(z - am) \end{cases} \tag{16}$$

where $c_1 = k/n$ and $c_2 = k/m$ are the regression coefficients it follows from the linearity of the regressions that the correlation coefficient is

$$\rho = \sqrt{c_1c_2} = k/\sqrt{mn}.$$

If $m = n$, we have a well known result which is sometimes stated as follows: If y and z are affected by n equally likely causes of which k are common to both, then the correlation coefficient between y and z is equal to k/n.

NORTHWESTERN UNIVERSITY.

A NOTE ON CONFIDENCE INTERVALS AND INVERSE PROBABILITY

By Albert Wertheimer

The object of this note is to discuss a certain property of confidence intervals from the point of view of inverse probability. We shall not go into detailed applications, but merely into fundamental ideas, so we shall work with distribution functions that are continuous and satisfy conditions which are sufficient to insure the validity of the mathematical steps used.

A clear and concise statement of the subject is given in a paper by Neyman,[1] and we shall use it as the basis for our discussion. His presentation can be summarized as follows: Let x be a sample statistic having a distribution function

$$p(x, \theta) \qquad \begin{array}{c} x_1 \leq x \leq x_2 \\ \theta_1 \leq \theta \leq \theta_2 \end{array}$$

where θ is a parameter of the population. Now define two monotonic functions

$$x = f(\theta); \qquad x = g(\theta) \qquad \begin{array}{c} x_1 \leq x \leq x_2 \\ \theta_1 \leq \theta \leq \theta_2 \end{array}$$

such that $f(\theta) < g(\theta)$, and

$$\int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon, \qquad \text{for all } \theta. \tag{1}$$

Let the prior distribution function of θ be

$$\psi(\theta) \qquad \theta_1 \leq \theta \leq \theta_2 .$$

It then follows directly that the probability for any pair of values (x, θ) lying within the region enclosed by the curves is given by

$$\int_{\theta_1}^{\theta_2} \psi(\theta)\, d\theta \int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon. \tag{2}$$

regardless of the prior function $\psi(\theta)$. His conclusion then is this: Stating that

$$g^{-1}(x) \leq \theta \leq f^{-1}(x) \tag{3}$$

every time the observation gives us a value of x equal to that given in (3) we may in any one instance be wrong; this will happen only if the pair (x, θ) for this observation lies outside the region enclosed by the curves; but from (2) the probability for this to happen is ϵ. This statement is equivalent to saying that

[1] *Journal of the Royal Statistical Society*, Vol. 97, part IV, 1934; pp. 589–93.

if for every observed x we write the inequality (3), then for a large number of samples, the fraction $1 - \epsilon$ of the inequalities will be found correct.

We note here that this is true only if in the inequality (3) x is presumed to range over its entire interval of definition. But if for an observation $x = x'$, we mean to consider the corresponding inequality

$$g^{-1}(x') \leq \theta \leq f^{-1}(x') \tag{4}$$

as one member of the class of inequalities that could be written just for those samples that had $x = x'$, then we can not assert that the inequality (4) has a probability of $1 - \epsilon$ of being correct. In fact, any probability statement dealing with this class must involve the prior distribution function $\psi(\theta)$; and if it is not given, then we do not know in what percent of cases the restricted inequality (4) will be found correct.

Let us nevertheless approach the problem from the viewpoint of inverse probability. Having observed $x = x'$, the posterior probability of inequality (4) being correct is

$$\eta(x') = \frac{\int_{g^{-1}(x')}^{f^{-1}(x')} \psi(\theta)p(x', \theta)\, d\theta}{\int_{\theta_1}^{\theta_2} \psi(\theta)p(x', \theta)\, d\theta} \tag{5}$$

the numerator being the probability for the simultaneous occurrence of

$$x = x'; \qquad g^{-1}(x') \leq \theta \leq f^{-1}(x'),$$

and the denominator the probability[2] that $x = x'$, θ lying anywhere between θ_1 and θ_2.

As long as $\psi(\theta)$ is unknown $\eta(x')$ cannot be evaluated; however its average value $\bar{\eta}(x)$ with respect to x can be evaluated. By definition of an average,

$$\bar{\eta}(x) = \int_{x_1}^{x_2} \eta(x)\, dx \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta \tag{6}$$

From (5) we have

$$\int_{g^{-1}(x)}^{f^{-1}(x)} \psi(\theta)p(x, \theta)\, d\theta = \eta(x) \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta \tag{7}$$

Integrating both sides of (7) over the entire range of x we get

$$\begin{aligned}\int_{x_1}^{x_2} dx \int_{g^{-1}(x)}^{f^{-1}(x)} \psi(\theta)p(x, \theta)\, d\theta &= \int_{x_1}^{x_2} \eta(x)\, dx \int_{\theta_1}^{\theta_2} \psi(\theta)p(x, \theta)\, d\theta \\ &= \bar{\eta}(x)\end{aligned}$$

[2] When we say probability that $x = x'$, we mean the probability that x will lie in the internal $x \pm \frac{1}{2}dx$ to within terms of order dx.

Interchanging the order of integration, as is permissible under the assumptions, we get

$$\bar{\eta}(x) = \int_{\vartheta_1}^{\theta_2} \psi(\theta)\, d\theta \int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx$$

But since

$$\int_{f(\theta)}^{g(\theta)} p(x, \theta)\, dx = 1 - \epsilon, \qquad \text{for all } \theta$$

we finally get

$$\bar{\eta}(x) = 1 - \epsilon$$

Thus when approached from the standpoint of inverse probability we see that the average value of the posterior probability of the inequality (4) is precisely the quantity $1 - \epsilon$ regardless of the prior distribution function $\psi(\theta)$.

In conclusion it is a pleasure to thank Dr. Deming for the criticisms and suggestions which he has made in connection with this note.

Navy Department,
Washington.

NOTE ON A MATCHING PROBLEM

By Solomon Kullback

1. **Introduction.** There is to be found in the literature [1] a number of discussions of the matching problem i.e., the problem of deriving the distribution of the number of correct matchings when two sequences of elements are placed in correspondence. However, the formulation of the matching problem discussed and illustrated herein is somewhat different from those problems already discussed in the literature [1], and may be of interest. A rather general statement of the problem follows.

2. **The Problem.** Consider urns U_i, $i = 1, 2, \cdots, n$ each of which contains some or all of the r different elements $E_1, E_2, \cdots, E_r$. The relative proportions of the r elements in the i-th urn are $p_{i1}, p_{i2}, \cdots, p_{ir}$ $(i = 1, 2, \cdots, n)$ such that

$$p_{i1} + p_{i2} + \cdots + p_{ir} = 1 \qquad i = 1, 2, \cdots, n \tag{1}$$

$$p_{i1}^2 + p_{i2}^2 + \cdots + p_{ir}^2 = p_i \qquad i = 1, 2, \cdots, n \tag{2}$$

(Some p_{ij} $i = 1, 2, \cdots, n, j = 1, 2, \cdots, r$ may be zero).

Assuming each urn to be an infinite source, consider two sequences made by drawing, at random, a single element from each urn in turn. If the two sequences are placed in correspondence there will be a number of correct matchings. What is the distribution of the number of correct matchings if the foregoing process be indefinitely repeated?

3. **Solution of the Problem.** The probability that the elements in the k-th position of the two sequences match may be derived by the following simple considerations. Since all the drawings are independent, the probability that both elements in the k-th position are E_m is p_{km}^2. Accordingly, the probability that both elements are the same, irrespective of their particular identity is $p_{k1}^2 + p_{k2}^2 + \cdots + p_{kr}^2 = p_k$.

The theory for the number of correct matchings in this case thus corresponds to that for the Poisson series, which is well known [2]. For the special case in which $p_k = p$, $k = 1, 2, \cdots, n$ the distribution of the number of correct matchings is in accordance with the binomial $(q + p)^n$ where $q = 1 - p$.

4. **Numerical Illustration and Verification.** The following illustration corresponds to the special case in which the urns are taken to be identical with equal proportions of each of the r elements.

Random sequences of 300 digits each were matched and the number of correct matchings recorded. The result of 457 such observations is given in Table 1.

TABLE 1

Observed distribution of number of correct matchings per sequences of 300 random digits each

Number of correct matchings	Observed frequency	Number of correct matchings	Observed frequency
18	1	32	35
19	2	33	25
20	3	34	15
21	5	35	20
22	9	36	20
23	22	37	17
24	18	38	6
25	21	39	10
26	41	40	7
27	28	41	1
28	30	42	3
29	31	43	1
30	42	44	0
31	42	45	2
		Total	457

Average number of correct matchings	Standard deviation
29.9934	4.8484

TABLE 2

Values of $P_x = (300!/x!(300-x)!)(0.1)^x(0.9)^{300-x}$

x	P_x	x	P_x	x	P_x	x	P_x
14	0.00033	23	0.03240	32	0.06920	41	0.00875
15	.00070	24	.04156	33	.06245	42	.00599
16	.00139	25	.05099	34	.05499	43	.00400
17	.00257	26	.05992	35	.04601	44	.00259
18	.00449	27	.06756	36	.03763	45	.00164
19	.00741	28	.07319	37	.02984	46	.00101
20	.01156	29	.07628	38	.02294	47	.00061
21	.01713	30	.07656	39	.01713	48	.00036
22	.02413	31	.07409	40	.01242	49	.00020

In accordance with paragraph 3, the distribution in Table 1 should correspond to the binomial distribution with $n = 300$ and $p = 10(1/10^2) = 1/10$. For the

TABLE 3

Comparison of observed distribution with the theoretical distribution
457 (0.9 + 0.1)^{300}

Number of correct matchings	Frequency		
	Observed	Theoretical	
	F_0	f	$F = 457f$
14–16	0	0.00242	1.1
17–19	3	.01447	6.6
20–22	17	.05282	24.1
23–25	61	.12495	57.1
26–28	99	.20067	91.7
29–31	115	.22693	103.7
32–34	75	.18614	85.1
35–37	57	.11348	51.9
38–40	23	.05249	24.0
41–43	5	.01874	8.6
44–46	2	.00524	2.3
47–49	0	.00117	.5
	457		456.7

TABLE 4

F_0	F	$(F_0 - F)^2/F$	
0, 3	1.1, 6.6	2.87	
17	24.1	2.09	
61	57.1	.27	
99	91.7	.58	$\chi_0^2 = 10.48$
115	103.7	1.23	
75	85.1	1.20	
57	51.9	.50	$n = 8$
23	24.0	.04	
5, 2, 0	8.6, 2.3, .5	1.70	$P(\chi^2 > \chi_0^2) = .235$
		10.48	

binomial distribution we have $m = np = 30$, $\sigma = \sqrt{npq} = \sqrt{27} = 5.1962$. To compare the observed distribution with the expected distribution we calcu-

lated the values of $P_x = (300!/x!(300 - x)!)(0.1)^x(0.9)^{300-x}$ for values of x from 14 to 49 inclusive which are given in Table 2.

To compare the observed and the theoretical distributions, and test the "Goodness of Fit," the distributions were grouped in classes of three. The results are shown in Tables 3 and 4.

5. **Conclusion.** The agreement between the observed distribution and the theoretical distribution derived on the basis of the argument in paragraph 3 is quite satisfactory.

We have shown herein, that if two sequences be matched under certain conditions, the distribution of the number of correct matchings will, in general, be that of a Poisson series and in special cases the binomial distribution. The theory was illustrated by an experiment which yielded results in satisfactory agreement with the theory.

REFERENCES

[1] An indication of, and some of the problems discussed, will be found in: E. G. OLDS, "A moment generating function which is useful in solving certain matching problems," *Bull. Am. Math. Soc.*, Vol. 44, June 1938, pp. 407–413.

[2] Some references to discussions of the theory, moments, and distribution connected with the Poisson series follow:

(a) H. L. RIETZ, *Mathematical Statistics*, Carus Mathematical Monograph No. 3, Open Court Pub. Co. (1927), pp. 148–152.

(b) CHARLES JORDAN, *Statistique Mathematique*, pp. 109–110.

(c) T. KAMEDA, "Theory of generating functions and its application to the theory of probability," *Journal of the Faculty of Science*, Imperial Univ. Tokyo, Section I, Vol. I, Part 1, (1925), Theorem XXIV, p. 48.

(d) P. R. RIDER, "The third and fourth moments of the generalized Lexis Theory," *Metron*, Vol. 12, No. 1, (1934) p. 195.

(e) S. KULLBACK, On the Bernoulli Distribution, *Bull. Am. Math. Soc.*, Vol. 41, (1935) p. 861.

THE GEORGE WASHINGTON UNIVERSITY.

REPORT OF THE ANNUAL MEETING OF THE INSTITUTE

The fourth annual meeting of the Institute of Mathematical Statistics was held in Detroit, Michigan, on December 27–29, 1938, in conjunction with the meetings of the American Statistical Association and the Econometric Society. The program for the meetings was arranged by Professors S. S. Wilks and B. H. Camp.

On Tuesday morning, December 27, the Institute held a session devoted to contributed papers with Professor B. H. Camp, president of the Institute in the chair. At that time the following papers were presented:

1. *Generalizations of the Laplace-Liapounoff Theorem.*
   W. G. Madow, Millbank Management Corporation, New York.
2. *The standard errors of the geometric and harmonic means.*
   Nilan Norris, Hunter College.
3. *Note on an integral equation in population analysis.*
   Alfred J. Lotka, Metropolitan Life Insurance Company, New York.
4. *Optimum fiducial regions for simultaneous estimation of several population parameters from large samples.*
   S. S. Wilks, Princeton University.
5. *A mathematical contribution to immigration assessment.*
   Churchill Eisenhart, University of Wisconsin.
6. *Contributions to the theory of statistical estimation.*
   A. Wald, Columbia University.
7. *On the hypotheses underlying the applications of statistical methods to routine laboratory analyses.*
   J. Neyman, University of California.
8. *Commodity transformations and matrices.*
   Harold Hotelling, Columbia University.
9. *Remarks on two methods of sample inspection.*
   E. G. Olds, Carnegie Institute of Technology.

Abstracts of these papers are given at the close of this report.

Immediately following the session just described, the Institute convened in business session. At that time President Camp announced that the newly elected officers for the year 1939 are: President, P. R. Rider, Washington University; Vice-Presidents, C. C. Craig, University of Michigan, and S. S. Wilks, Princeton University; Secretary-Treasurer, A. T. Craig, University of Iowa.

The annual luncheon of the Institute was held at one o'clock on the same day. At the luncheon, Dr. Walter A. Shewhart, of the Bell Telephone Laboratories addressed the Institute on "The Future of Statistics in Mass Produc-

tion." A summary of this address is included among the abstracts at the close of this report.

On Wednesday morning, December 28, the Institute and the Statistical Association held a joint session devoted to the teaching of Business Statistics. Professor T. H. Brown presided. The following papers constituted the program:

1. *The teaching of undergraduate students.*
   L. S. Kellogg, Ohio State University.
2. *The teaching of graduate students.*
   O. W. Blackett, University of Michigan.
3. *A bead-sampling machine for use in the class room.*
   Dickson H. Leavens, Cowles Commission for Research in Economics.

Discussion: Harry P. Hartkemeier, University of Missouri.
Richard L. Kozelka, University of Minnesota.

On the afternoon of the same day, the Biometric Section of the Statistical Association and the Institute presented the following program on Statistical Methods in Genetics Problems with Professor Lowell J. Reed as chairman:

1. *Tests of simple Mendelian inheritance in randomly collected data of one and two generations.*
   Laurence H. Snyder and Charles W. Cotterman, Ohio State University.
2. *Statistical studies of the familial aspects of cancer in humans.*
   Herbert L. Lombard, Massachusetts State Department of Public Health.
3. *Application of the method of likelihood ratios to the testing of hypotheses of simple Mendelian inheritance.*
   S. S. Wilks, Princeton University.
4. *The application of statistical techniques to egg production data for the formulation of a breeding program.*
   W. C. Thompson, New Jersey Agricultural Experiment Station.

The Program Committees of the Institute and the Statistical Association arranged a joint session on Representative Sampling for Thursday afternoon, December 29. At that time the following papers were presented, with Professor Harold Hotelling presiding:

1. *On the mathematics of the representative method.*
   Allen T. Craig, University of Iowa.
2. *Application of the theory of sampling to large scale surveys and censuses.*
   Frederick F. Stephan, American Statistical Association.
3. *Further remarks on the mathematical aspects of representative sampling.*
   J. Neyman, University of California.

Discussion: Samuel A. Stouffer, University of Chicago.
Churchill Eisenhart, University of Wisconsin.
P. J. Rulon, Harvard University.

The final session of the meetings was held on Thursday evening. This was a joint session with the Econometric Society and was devoted to Mathematical

Statistics in Economics. Professor Irving Fisher presided and the following papers were given:

1. *On the hypothesis of linearity of regression in economic research.*
   J. Neyman, University of California.
2. *The selection of variates for use in prediction.*
   Harold Hotelling, Columbia University.
3 *Decomposition of time series on the basis of non-correlation principle.*
   Wassily Leontief, Harvard University.

Discussion: William G. Madow, Millbank Management Corporation, New York.
Gerhard Tintner, Iowa State College.

A. T. Craig, *Secretary.*

ABSTRACTS OF PAPERS

(Presented on December 27, 1938, at the Detroit meeting of the Institute)

Generalizations of the Laplace-Liapounoff Theorem. W. G. MADOW, Milbank Management Corporation, New York.

The Laplace-Liapounoff Theorem states conditions under which a linear function of chance variables has a normal limiting distribution.

In dealing with limiting distributions arising in the analysis of variance, regression analysis, etc., there occurred problems which required for their solution the derivation of the joint limiting distribution of several linear functions of chance variables and the joint limiting distribution of functions which were linear in one set of chance variables for fixed values of other sets of chance variables.

These problems were solved by a matrix formulation of the Laplace-Liapounoff Theorem and by the introduction of a function whose convergence to zero in probability provided a sufficient condition for the existence of normal limiting distributions.

Various generalizations with a view towards applications in multi-variate statistical analyses are discussed. The theorems provide a rigorous and complete basis for the derivation of limiting distributions of quadratic and bilinear forms.

The Standard Errors of the Geometric and Harmonic Means. NILAN NORRIS, Hunter College.

Although certain properties of the geometric and harmonic means have been investigated extensively, there seems to have been no derivation of expressions for their variances in cases where they are used as estimates of parameters of parent populations.

Application of the modern theory of estimation makes it possible to develop simple and useful formulae for the standard errors of these two averages for each of the respective general classes of cases in which they are most suitable.

As in other instances in which standard errors are used in tests of significance, fiducial or confidence limits may be employed to overcome certain limitations of the outmoded practice of relying solely on multiples of either probable or standard errors to determine whether or not a result exists merely because of sampling fluctuations.

Note on an Integral Equation in Population Analysis. ALFRED J. LOTKA, Metropolitan Life Insurance Company, New York.

In a population in which immigration and emigration are negligible, the number $N(t)$ of the population at time t is connected with the annual births $B(t)$ and the probability $p(a)$ of surviving from birth to age a, by the obvious relation

$$N(t) = \int_0^\infty B(t - a)p(a)\, da. \tag{1}$$

If $B(t)$ and $p(a)$ are given, $N(t)$ follows at once by direct integration. The inverse problem, given $N(t)$, to find $B(t)$, requires separate treatment. The case that $N(t)$ is given or can be expressed as a sum of exponential functions has been discussed by the

author on a former occasion. In the present communication it is shown how the function $B(t)$ can be expressed as a series proceeding in ascending derivatives of $N(t)$.

A second solution is also offered in which

$$\frac{B(t)}{N(t)} = b(t),$$

the birth rate per head is obtained as a series, the first and dominating term of which is

$$b(t) = \frac{1}{\int_0^\infty e^{-r_t a} p(a)\, da} \tag{2}$$

where r_t is the rate of natural increase at time t. This development is of interest because it corresponds to the expression for b in a population with constant birth rate, death rate, and rate of natural increase; that is

$$b = \frac{1}{\int_0^\infty e^{-ra} p(a)\, da} \tag{3}$$

so that the new expression represents $b(t)$ as the corresponding value of b in a Malthusian population, plus a series of correcting terms.

Optimum Fiducial Regions for Simultaneous Estimation of Several Population Parameters from Large Samples. S. S. WILKS, Princeton University.

If a population has a distribution law $f(x, \theta)$ where x is the variate and θ is a parameter, it is known (*Annals of Mathematical Statistics*, Vol. IX (1938) pp. 166–175) that under rather general conditions, confidence intervals, for a given confidence coefficient α, which are shortest on the average, can be obtained from large samples of n items by solving the equations

$$\frac{\frac{\partial L}{\partial \theta}}{\sqrt{n}\sqrt{E\left[\frac{\partial \log f}{\partial \theta}\right]^2}} = \pm d_\alpha \tag{1}$$

for θ, where d_α is the normal deviate given by $\frac{1}{\sqrt{2\pi}}\int_{-d_\alpha}^{+d_\alpha} e^{-\frac{1}{2}t^2}\, dt = \alpha$. L is the logarithm of the likelihood, i.e. $L = \sum_{i=1}^{n} \log f(x_i, \theta)$, where E denotes mean value with respect to the probability law $f(x, \theta)$.

The present paper is an extension of the foregoing results to the case of several parameters. It is shown under fairly general conditions that if the distribution law of x if a function $f(x, \theta_1, \cdots, \theta_h)$ depending on h parameters, then for a confidence coefficient α the fiducial region of the θ's which is smallest in size *on the average* is given by the region in the space of the θ's for which

$$\frac{1}{n}\sum_{i,j=1}^{h} a_{ij}\left(\frac{\partial L}{\partial \theta_i}\right)\left(\frac{\partial L}{\partial \theta_j}\right) \leq \chi_\alpha^2 \tag{2}$$

where χ^2_α is such that $P(\chi^2 \leq \chi^2_\alpha) = \alpha$, where the probability is calculated from a χ^2 distribution with h degrees of freedom. The matrix $\| a_{ij} \|$ is the inverse of the matrix whose general element is

$$E\left[\frac{\partial \log f}{\partial \theta_i} \cdot \frac{\partial \log f}{\partial \theta_j}\right]$$

Similar results hold when f is a function of several random variables as well as the parameters $\theta_1, \theta_2, \cdots, \theta_h$.

A Mathematical Contribution to Immigration Assessment. CHURCHILL EISENHART, University of Wisconsin.

A certain problem is assessing the size of an immigration can be stated mathematically as follows: A sample of size N is drawn at random from a population in which the probability of A is p. Let "not-A" be denoted by B. Then the sample will contain a frequency, say x, of A and $N - x$ of B, x being a random variable. This sample is now mixed together with a very much larger sample in which the elements are all B's, and the B's belonging to the original sample lost sight of. The problem is to estimate N from the observed frequency of A, namely x, in the composite sample, p being assumed known. The maximum likelihood estimate of N is x/p. For large values of x, and a fortiori of N, confidence intervals for N take the form $N_1 \leq N \leq N_2$ where

$$N_1 = \frac{\{\sqrt{q^2t^2 + 4q(x - \frac{1}{2})} - qt\}^2}{4pq},$$

$$N_2 = \frac{\{\sqrt{q^2t^2 + 4q(x + \frac{1}{2})} + qt\}^2}{4pq},$$

$$q = 1 - p,$$

and the confidence coefficient is .95 if t is set equal to 1.96, and is .99 if t is set equal to 2.58. For small values of x the solution is more difficult but charts are being prepared from which the confidence intervals can be read off.

A Contribution to the Theory of Statistical Estimation. A. WALD, Columbia University.

Let us denote by $f(x, \theta)$ a probability density function, where θ is a parameter. Denote by Ω the set of all possible values of θ. The assumption that θ belongs to a subset ω of Ω is called a hypothesis. Let us consider a system S of subsets of Ω. Denote the hypothesis corresponding to an element ω of S by H_ω and the system of all hypotheses corresponding to the elements of S by H_S. Denote by E a sample point in the n-dimensional sample space drawn from a population with the probability density function $f(x, \theta)$, where the value of θ is unknown. We have to decide by means of the sample point E which hypothesis of the system H_S should be accepted. That is to say, for each hypothesis H_ω we have to choose a region of acceptance M_ω in the sample space. The hypothesis H_ω will be accepted if and only if the sample point E falls in the region M_ω. Denote by M_S the system of all regions M_ω. The statistical problem to be solved is the question of *how the system of regions M_S should be chosen?*

In order to answer this question, a non-negative weight function $w(\theta, \omega)$ is introduced, which is defined for all values θ and for all elements ω of S. The weight $w(\theta, \omega)$ expresses the loss caused by accepting H_ω if θ is true. The probability of accepting H_ω multiplied by the weight $w(\theta, \omega)$ is called the risk of accepting H_ω if θ is true. Denote this risk by

$r(\theta, H_\omega, M_S)$ (the risk depends obviously also on the system of regions M_S). The total risk of accepting a false hypothesis if θ is true, is given by

$$r(\theta, M_S) = \sum_\omega r(\theta, H_\omega, M_S)$$

where the summation is to be taken over all elements ω of S which do not contain θ.

Denote by $r(M_S)$ the maximum of $r(\theta, M_S)$ with respect to θ. The system M_S of regions for which $r(M_S)$ becomes a minimum is called the "best" system of regions relative to the weight function $w(\theta, \omega)$. Some properties of the best system of regions have been studied and the problem of its calculation has been treated.

On the Hypotheses underlying the Applications of Statistical Methods to Routine Laboratory Analyses. J. Neyman, University of California.

The problem considered is that of estimating the proportion, p, of certain designated elements of the population sampled, the estimate to be based on a random sample of n, drawn by some mechanical device, such as is sometimes used in industry and in laboratory work. Examples: (1) to estimate the proportion of defective manufactured products in mass production; (2) to estimate the proportion of seeds which are able to germinate in given conditions. One would expect that the sample proportion, say q, will be distributed in repeated samples according to the Binomial Law and that, consequently, in order to obtain the confidence limits for p, one should use the Clopper-Pearson graphs. However, the evidence obtained from special analyses on seed germination, made in the Seed Testing Station of Warsaw, Poland, shows that this assumption may not be true. The sampling there was carried out by means of a machine and involved a certain amount of mixing. As a result q was more stable than it was expected. It did not follow the Binomial Law at all, but a Normal one about p, with a standard deviation, σ, which could be well estimated from the sum of squares of deviations from respective means. For a considerable period of time (18 months) σ retained a constant value (a characteristic of the action of the sampling machine) which was rather smaller than $(n^{-1}q(1 - q))^{\frac{1}{2}}$.

Consequently, to have a preassigned frequency of correct statements concerning p, it was necessary to calculate the confidence intervals according to the formulae of the Normal Theory

$$q - t\sigma < p < q + t\sigma$$

with an appropriate value of t. Probably similar situations are rather common.

Remarks on Two Methods of Sampling Inspection. E. G. Olds, Carnegie Institute of Technology.

When the instructions for inspecting lots of size m specify that samples of size n be taken and the lot be passed without detailed inspection if no defectives are found, then, on the average, the maximum number of defectives are passed when the number of defects per lot is $\frac{m+1}{n+1}$ or $\frac{m+1}{n+1} - 1$.

If the quality of a lot is to be checked by drawing pieces until a fixed number of defective pieces are found, it is important to know that the expected number n_i necessary to obtain i defects is $i\frac{m+1}{p+1}$, where there are p defects in the lot. If $\frac{n_i}{i(m+1)}$ is used as an estimate of $\frac{1}{p+1}$, it is convenient to observe that the variance of $\frac{n_i}{i(m+1)}$ is

$$\frac{1}{p+2}\left[\frac{1}{p+1} - \frac{1}{m+1}\right]\left[\frac{1}{i} - \frac{1}{p+1}\right].$$

Commodity Transformations and Matrices. HAROLD HOTELLING, Columbia University.

If we regard the prices and quantities of n commodities as vectors we may apply the theory of linear transformations in various ways having economic and statistical significance. An example is the mixing of grains to produce results conforming to new specifications, as in international trade. Another kind of example arises in problems of multivariate statistical analysis such as those treated in my paper on "Relations between Two Sets of Variates" (*Biometrika*, 1936), concerned with properties invariant under internal linear transformations of the variates of each set. Prices transform contragrediently to quantities in all cases. Hence, if the prices and quantities of the same set of commodities are the two sets of variates, the allowable transformations are restricted. Consequently there are invariants in this case additional to those discussed in the paper mentioned. Another problem is the reduction of sets of linear demand functions to normal form and determination of invariants when transformations of prices must be contragredient to those of commodities. The question whether the demand functions are symmetrical is here of paramount importance, since symmetry is preserved by such transformations, and since there are known theoretical reasons to expect symmetry. For a non-singular set of linear symmetrical demand *or* supply functions there are no invariants under arbitrary sets of contragredient transformations; but for pairs of such sets of demand *and* supply functions there are invariants, namely the elementary divisors of the pair of matrices of coefficients. A set of demand or supply functions alone has invariants if its matrix B is not symmetrical. If B' denote the transverse or conjugate of B, the elementary divisors of $B + \lambda B'$ are such invariants.

The Future of Statistics in Mass Production.[1] WALTER A. SHEWHART, Bell Telephone Laboratories, New York.

Much has been written about the application of statistical theory and technique in studying, discovering, and measuring the effects of an existing system of unknown or chance causes. Much remains to be written about the application of statistical theory and technique in finding out how to tinker with and modify an existing chance cause system until it behaves as we would have it do. In research, we use statistical theory in helping to predict the future effects of some existing cause system. The statistician knows that his predictions will be valid if certain assumptions about the cause system are justified. Perhaps the most important assumption of this type is that the particular effects of a chance cause system under study are random. In mass production, however, the statistician has learned by experience that chance cause systems producing random effects don't just happen even under what we customarily consider to be the best regulated laboratory conditions. If the industrial statistician chooses to ignore this fact and makes predictions as if he were dealing with random cause systems, he may expect many of his predictions to fall far short of the truth: what is more, he knows that this fact will be discovered and his work discredited because in a continuing mass production process predictions are sure to be checked. Hence the industrial statistician in mass production must start not where the research statistician leaves off but, as it were, before the research statistician begins: that is, he must start by developing techniques for determining when we are justified in assuming that the underlying cause system is random. We thus arrive at a good starting point from which to consider the future of statistics in mass production.

Experience in the control of quality has provided a practical technique for detecting

[1] Summary of an address delivered at a luncheon meeting of the Institute of Mathematical Statistics, Detroit, Michigan, December 27, 1938.

and eliminating assignable causes of variability in the production process until a state of statistical control is reached where predictions based upon the assumption of randomness are likely to prove valid. It has also been shown elsewhere that by elimination of assignable causes of variability, we may make the most efficient use of raw materials, maximize the assurance of maintaining standard quality of manufactured goods, minimize the cost of inspection, and minimize the cost of rejections. Hence we may conclude that the use of statistics in mass production can be made to pay good dividends: such use can be made to have a bright future. On what then does that future depend?

To answer this question, we must consider the following three fundamental steps in the process of mass production:

I. The specification of the quality of the thing wanted.
II. The production of things designed to meet the specification.
III. The inspection of the things produced to see if they meet the specification.

An outstanding characteristic of the first step, specification, is the necessity of setting up and living within what we term a tolerance range[2] for each specified quality characteristic. If a producer contracts to live within some specified range and in taking steps II and III fails to do so, he usually loses a lot of money. Hence he must know how to set tolerance limits that he can meet. Moreover, if he is to be able to make the most efficient use of materials in many instances, he must close up as much as he feasibly can on the specified tolerance range.

Obviously, however, one can not specify a practically attainable tolerance range out of thin air: one must be limited by what it is possible to do under commercial conditions of production in step II and this in turn is revealed by inspection in step III. We must also take into account the fact previously noted that any manufacturing process to begin with is almost certain not to be in a state of statistical control. In fact, this state can only be approached through the application of certain statistical techniques that have been found useful in detecting the presence of assignable causes that can be found and removed. A point to be stressed is that the three steps—specification, production, and inspection—in mass production, cannot be taken independently: instead, they must be coördinated. It also may be shown that maximum effectiveness in the use of statistical theory can only be attained by coördinating the applications in each of the three steps. It is significant to note that in order to attain the most efficient use of materials and processes by minimizing the tolerance range and in order to minimize the cost of production, one must make effective use of the results obtained in the course of commercial production, particularly those in the third step, inspection. In fact, the three steps might be thought of as constituting a scientific experiment in which the objective is the attainment of the most efficient use of available materials in the production of manufactured goods.

Broadly speaking, the statistician of the future has before him the opportunity of helping to develop this fundamental type of experiment in many respects like the way he is successfully helping today in so many fields of research to design experimental procedures that make the most efficient use of human effort. Certain differences, of course, exist. For example, as already noted, he must start by designing a statistical control technique for randomizing, as it were, the cause system through the elimination of assignable causes. Then he can use modern statistical techniques of research in much the same way described in the literature with reasonable assurance that resulting predictions will be found valid because he has first randomized his cause system. He must, however, go farther than indicated in the current literature of statistical research in that he must provide operationally verifiable meanings for statistical terms such as random variable, accuracy, pre-

[2] The tolerance range is not to be confused with the fiducial range of modern statistics. The distinction between the two is set forth at some length in a forthcoming publication, *Statistical Method from the Viewpoint of Quality Control*, to be published shortly by the Graduate School of the United States Department of Agriculture.

cision, true value, probability, degree of rational belief, and the like.[3] This is particularly necessary in the steps of specification and inspection because the specification is often made the basis of a contractual agreement between producers and consumers.

There is a sense in which the statistician's problem in helping to develop the mass production process so as to make the most effective use of information yielded by the process is much more complicated than the design of experiment usually considered in the literature of statistics. Whereas the customary statistical theory of design of experiment in research is concerned with comparatively small-scale experiments carried out under controlled conditions of the laboratory by a few people, the corresponding development of the mass production process must be carried out under commercial conditions on a large scale involving large numbers of people. For example, the three steps in the mass production process are usually carried out either by different companies or by different departments of the same company. The steps may involve the coördinated effort of literally hundreds and even thousands of employees, including physicists, chemists, engineers, sales agents, purchasing agents, lawyers, and economists. Very few of these people have ever had any training in statistics or probability and yet many of them must be sold on the use of statistics if the statistician is to have an opportunity of making his full contribution to the social and economic effectiveness of the mass production process. This situation constitutes a problem not only for those now in industry but also for those responsible for training the industrial leaders of tomorrow so that they will have sufficient knowledge of statistics to help them recognize the potential contributions of statistical theory and technique.

In conclusion, then, we may say that in the future the statistician in mass production must do more than simply study, discover, and measure the effects of existing chance cause systems: he must devise means for modifying these cause systems in the best way to satisfy human wants. The statistician in mass production must not be satisfied with simply measuring demand for goods; he must help change that demand by showing, among other things, how to close up the tolerance range and improve the quality of goods. He must not be content with measuring production costs; he must help decrease production costs through the use of the techniques of statistical control.

The future contribution of statistics in mass production lies not so much in solving the problems usually put to the statistician today by those not statistically trained as in taking a hand in helping to coördinate the steps of specification, production, and inspection considered as a scientific experiment for making the most efficient use of human effort in the production of goods to satisfy human wants. The long range contribution of statistics depends perhaps not so much upon getting a lot of highly trained statisticians into industry in the immediate future as it does in creating a statistically minded new generation of those physicists, chemists, engineers, and others who will in any way have a hand in developing and directing the mass production process of tomorrow.

---

[3] An initial step in this direction has been taken in my Washington lectures. Loc. cit.

CONSTITUTION
OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE I

Name and Purpose

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

Membership

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.
2. Voting members of the Institute shall be (a) the Fellows, and (b) all others who have been members for twenty-three months prior to the date of voting.

ARTICLE III

Officers, Board of Directors, Committee on Membership, and Committee on Publications

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer, elected for a term of one year by a majority ballot at the annual meeting of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers and the previous President.
3. The Institute shall have a Committee on Membership composed of three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.
4. The Institute shall have a Committee on Publications composed of three Members or Fellows elected by the Board of Directors. The President shall designate a Vice-President as Ex Officio Chairman of this Committee.

ARTICLE IV

Meetings

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such

time as the Board of Directors may designate. Additional meetings may be called from time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute.

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the Board by the Secretary-Treasurer at least five days prior to the date set therefor. Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3. The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting.

4. At a regularly convened meeting of the Board of Directors, three members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

ARTICLE V

Publications

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. Other publications may be originated by the Board of Directors as occasion arises.

ARTICLE VI

Expulsion or Suspension

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

ARTICLE VII

Amendments

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

BY-LAWS

ARTICLE I

Duties of the Officers, Board of Directors, Committee on Membership, and Committee on Publications

1. The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present, shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be submitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2. The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute. He shall send out calls for annual dues and acknowledge receipt of same; pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditor shall report to the President.

3. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute.

4. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership.

5. The Committee on Publications, under the general supervision of the Board of Directors, shall have charge of all matters connected with the publications of the Institute, and of all books, pamphlets, manuscripts and other literary or scientific material collected by the Institute. Once a year this Committee shall cause to be printed in the Official Journal the Constitution and By-Laws and a classified list of all the Members and Fellows of the Institute.

ARTICLE II

Dues

1. Members shall pay five dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members shall pay

five dollars annual dues. The annual dues of Fellows shall be five dollars. The annual dues of Sustaining Members shall be fifty dollars. Honorary Members shall be exempt from all dues.

2. Annual dues shall be payable on the first day of January of each year.

3. The annual dues of a Fellow or Member include a subscription to the Official Journal. The annual dues of a Sustaining Member include two subscriptions to the Official Journal.

4. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues may be six months in arrears, and to accompany such notice by a copy of this Article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent one to the Board of Directors, by whom the person's name may be stricken from the rolls and all privileges of membership withdrawn. Such person may, however, be re-instated by the Board of Directors upon payment of the arrears of dues.

ARTICLE III

Salaries

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee.

ARTICLE IV

Amendments

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors.

DIRECTORY OF INSTITUTE OF MATHEMATICAL STATISTICS

(As of January 1, 1939)

Please notify the Secretary, A. T. Craig, University of Iowa, Iowa City, Iowa, of any changes of addresses or errors in names, titles or addresses as printed.

Acerboni, Dr. Argentino V., Larroque 232 Banfield, Buenos Aires, Republic of Argentina, South America.

Alter, Professor Dinsmore, Director of Griffith Observatory, Los Angeles, California.

Arnold, Professor Herbert E., Department of Mathematics, Wesleyan University, Middletown, Connecticut.

Aroian, Professor Leo, 522 Hill Street, Ann Arbor, Michigan.

Bacon, Professor H. M., Box 1144, Stanford, California.

Baker, Dr. G. A., Experiment Station, College of Agriculture, University of California, Davis, California..

Barral-Souto, Dr. José, Córdoba 1459, Buenos Aires, Republic of Argentina, South America.

Beall, Mr. Geoffrey, 416 Queen Street, Chatham, Ontario, Canada.

Bennett, Professor A. A., Department of Mathematics, Brown University, Providence, Rhode Island.

Berkson, Dr. Joseph, The Mayo Clinic, Rochester, Minnesota.

Bernstein, Professor Felix, Department of Biometry, New York University, New York, New York..

Brooks, Mr. A. G., Western Electric Company, Hawthorne Station, Chicago, Illinois.

Burgess, Dr. R. W., Western Electric Co., 195 Broadway, New York, New York.

Burr, Professor Irving W., Antioch College, Yellow Springs, Ohio.

Bushey, Professor J. Hobart, Department of Mathematics, Hunter College, Park Avenue and 68th Street, New York, New York.

Camp, Professor B. H., Department of Mathematics, Wesleyan University, Middletown, Connecticut..

Camp, Professor C. C., Department of Mathematics, University of Nebraska, Lincoln, Nebraska.

Carver, Professor H. C., Department of Mathematics, The University of Michigan, Ann Arbor, Michigan..

Chapman, Mr. Roy A., 400 Union Building, New Orleans, Louisiana.

Clark, Professor Andrew G., 1012 Laporte Avenue, Ft. Collins, Colorado.

Coleman, Mr. E. P., Department of Mathematics, The Municipal University of Omaha, Omaha, Nebraska.

Cox, Dr. Gerald J., The Mellon Institute, Pittsburgh, Pennsylvania.

Craig, Professor A. T., Department of Mathematics, The University of Iowa, Iowa City, Iowa..

Craig, Professor C. C., 3020 Angell Hall, The University of Michigan, Ann Arbor, Michigan..

Crathorne, Professor A. R., Department of Mathematics, The University of Illinois, Urbana, Illinois..

CROWE, PROFESSOR S. E., 137 University Drive, East Lansing, Michigan.
CURTISS, DR. J. H., Belle Ayre Apartments, Ithaca, New York.
DELURY, DR. D. B., Department of Mathematics, The University of Toronto, Toronto 5, Canada.
DEMING, DR. W. E., Fertilizer Investigations, Friendship Post Office, Washington, D. C..
DODD, PROFESSOR E. L., Department of Mathematics, The University of Texas, Austin, Texas..
DODGE, MR. HAROLD F., Bell Telephone Lab., 463 West Street, New York, New York.
DOOB, DR. J. L., Department of Mathematics, The University of Illinois, Urbana, Illinois..
DUNLAP, PROFESSOR JACK W., Catharine Strong Hall, University of Rochester, Rochester, New York.
DWYER, PROFESSOR PAUL S., 407 Camden Court, Ann Arbor, Michigan.
EDGETT, DR. G. L., Queen's University, Kingston, Ontario, Canada.
EISENHART, DR. CHURCHILL, Department of Mathematics, The University of Wisconsin, Madison, Wisconsin.
ELSTON, MR. JAMES S., The Travelers Insurance Company, Hartford, Connecticut.
ELVEBACK, MISS MARY L., Room 11, Library, University of Minnesota, Minneapolis, Minnesota.
EVANS, PROFESSOR H. P., North Hall, The University of Wisconsin, Madison, Wisconsin.
FEINLER, REV. F. J., 821 East C Street, Grants Pass, Oregon.
FELDMAN, DR. H. M., Soldan High School, St. Louis, Missouri.
FERTIG, DR. JOHN W., Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Maryland.
FISCHER, DR. C. H., Department of Mathematics, Wayne University, Detroit, Michigan.
FISHER, PROFESSOR IRVING, 460 Prospect Street, New Haven, Connecticut..
FOSTER, MR. RONALD M., 122 East Dudley Avenue, Westfield, New Jersey.
FRANKEL, MR. LESTER R., 1445 Otis Place NW, Washington, D. C.
FRY, DR. T. C., 463 West Street, New York, New York..
GAVETT, PROFESSOR G. I., Department of Mathematics, The University of Washington, Seattle, Washington.
GILL, MR. JOHN P., Department of Mathematics, The University of Alabama, University, Alabama.
GIRSHICK, MR. M. A., Bureau of Home Economics, Department of Agriculture, Washington, D. C.
GLOVER, PROFESSOR J. W., 620 Oxford Road, Ann Arbor, Michigan..
GREENWOOD, DR. J. A., Box 4723, Duke University, Durham, North Carolina.
GREVILLE, DR. T. N. E., Department of Mathematics, The University of Michigan, Ann Arbor, Michigan.
GROVE, PROFESSOR C. C., 143 Melburn Avenue, Baldwin, L. I., New York.
HART, PROFESSOR W. L., Department of Mathematics, The University of Minnesota, Minneapolis, Minnesota.
HENDERSON, MR. ROBERT, Crown Point, Essex County, New York..
HENDRICKS, MR. WALTER A., National Agricultural Research Center, Beltsville, Maryland.
HENRY, MR. M. H., 829 West Allegan Avenue, Lansing, Michigan.
HOEL, DR. PAUL G., Department of Mathematics, Oregon State College, Corvallis, Oregon.
HOTELLING, PROFESSOR HAROLD, Fayerweather Hall, Columbia University, New York, New York..
HUNTINGTON, PROFESSOR E. V., 48 Highland Street, Cambridge, Massachusetts..
INGRAHAM, PROFESSOR M. H., North Hall, The University of Wisconsin, Madison, Wisconsin..
JACKSON, PROFESSOR DUNHAM, 119 Folwell Hall, The University of Minnesota, Minneapolis, Minnesota..
JANKO, DR. JAROSLAV, Praha XII, Krkonošská 3, Czechoslovakia.

Juran, Mr. J. M., Western Electric Co., 195 Broadway, New York, New York.
Keeping, Professor E. S., University of Alberta, Edmonton, Alberta, Canada.
Kelley, Professor Truman L., Lawrence Hall, Cambridge, Massachusetts..
Kenney, Mr. John F., Department of Mathematics, Northwestern University, Evanston, Illinois.
Kiernan, Professor Charles J., 19 Looker Street, Hillside, New Jersey.
Knowler, Dr. Lloyd A., Department of Mathematics, Hunter College, New York, New York.
Kossack, Mr. Carl F., RFD #2, Ann Arbor, Michigan.
Kullback, Dr. Solomon, Office of Chief Signal Officer, Washington, D. C..
Kurtz, Dr. Albert K., 173 Cornwall Street, Hartford, Connecticut.
Larsen, Professor Harold D., Department of Mathematics, University of New Mexico, Albuquerque, New Mexico.
Leavens, Mr. Dickson H., 301 Mining Exchange Building, Colorado Springs, Colorado.
Lemme, Professor Maurice M., Department of Mathematics, University of Toledo, Toledo, Ohio.
Livers, Mr. Joe J., 723 Oakland Avenue, Ann Arbor, Michigan.
Lorge, Dr. Irving, 525 West 120 Street, New York, New York.
Lotka, Dr. A. J., One Madison Avenue, New York, New York..
Madow, Dr. William G., Millbank Management Corporation, Chrysler Building, New York, New York.
Malzberg, Dr. Benjamin, New York State Department of Mental Hygiene, Albany, New York.
Mauchly, Professor J. W., Department of Physics, Ursinus College, Collegeville, Pennsylvania.
Mode, Professor E. B., 688 Boylston Street, Boston, Massachusetts.
Molina, Mr. E. C., 463 West Street, New York, New York..
Mudgett, Professor Bruce D., Department of Economics, University of Minnesota, Minneapolis, Minnesota.
MacLean, Mr. M. C., Chief of Census Analysis, Dominion Bureau of Statistics, Ottawa, Canada.
McEwen, Professor G. F., Scripps Institution, La Jolla, California.
Nesbitt, Dr. C. J., Department of Mathematics, The University of Michigan, Ann Arbor, Michigan.
Neyman, Professor J., Department of Mathematics, The University of California, Berkeley, California..
Norris, Dr. Nilan, Department of Economics, Hunter College, New York, New York.
Olds, Professor E. G., 953 La Clair Avenue, Regent Square, Pittsburgh, Pennsylvania.
Ollivier, Dr. Arthur, Box 405, State College, Mississippi.
Olpin, Mr. A. R., Kendall Mills, Paw Creek, North Carolina.
O'Toole, Dr. A. L., 1425 E. 60th Street, Chicago, Illinois..
Ore, Professor Oystein, Department of Mathematics, Yale University, New Haven, Connecticut.
Parente, Mr. A. R., 126 Church Street, Hamden, Connecticut.
Pixley, Professor H. H., Department of Mathematics, Wayne University, Detroit, Michigan.
Pollard, Professor H. S., Department of Mathematics, Miami University, Oxford, Ohio.
Reed, Professor L. J., 615 North Wolfe Street, Baltimore, Maryland.
Rider, Professor P. R., Department of Mathematics, Washington University, St. Louis, Missouri..
Rietz, Professor H. L., Department of Mathematics, The University of Iowa, Iowa City, Iowa..
Romig, Mr. H. G., 463 West Street, New York, New York.

ROOS, DR. CHARLES, Care of Mercer-Allied Corporation, 420 Lexington Avenue, New York, New York..
RULON, PROFESSOR P. J., 4 Emerson Hall, Harvard University, Cambridge, Massachusetts.
SCARBOROUGH, PROFESSOR J. B., Post Office Box 332, Annapolis, Maryland.
SHEWHART, DR. W. A., 158 Lake Drive, Mountain Lakes, New Jersey..
SHOHAT, PROFESSOR J., Department of Mathematics, The University of Pennsylvania, Philadelphia, Pennsylvania..
SPIEGELMAN, MR. MORTIMER, 325 West 86th Street, New York, New York.
STEPHAN, MR. F. F., 1626 K Street NW, Washington, D. C.
STOUFFER, PROFESSOR S. A., Department of Sociology, University of Chicago, Chicago, Illinois.
SWANSON, MR. A. G., General Motors Institute, Flint, Michigan.
THOMPSON, DR. WILLIAM R., 27 Oakwood Street, Albany, New York.
TOOPS, PROFESSOR HERBERT A., The Ohio State University, Columbus, Ohio.
TORREY, MISS MARY N., 463 West Street, New York, New York.
TRELOAR, PROFESSOR ALAN E., Department of Biometry, The University of Minnesota, Minneapolis, Minnesota.
TRUBRIDGE, DR. G. F. P., 98 Kingsway, Petts Wood, Kent, England.
VATNSDAL, MR. J. R., 304 Columbia Street, Pullman, Washington.
VICKERY, DR. C. W., 304 E. Fifth Street, Austin, Texas.
WALKER, PROFESSOR HELEN M., Teachers College, Columbia University, New York, New York.
WEI, MR. DZUNG-SHU, 19 Evans Street, Iowa City, Iowa.
WEIDA, PROFESSOR F. M., Department of Mathematics, The George Washington University, Washington, D. C..
WESCOTT, MR. MASON E., 1936 Greenwood Avenue, Wilmette, Illinois.
WELKER, MR. E. L., 160 Mathematics Building, University of Illinois, Urbana, Illinois.
WHITE, PROFESSOR A. E., Department of Mathematics, Kansas State College, Manhattan, Kansas.
WILKS, PROFESSOR S. S., Fine Hall, Princeton University, Princeton, New Jersey..
WILSON, DR. ELIZABETH W., One Waterman Street, Cambridge, Massachusetts.
WILSON, MR. W. P., Traffic Engineering Bureau, 206 Police Headquarters, City of Detroit, Detroit, Michigan.
WRIGHT, PROFESSOR SEWALL, Department of Zoology, University of Chicago, Chicago, Illinois..
WYCKOFF, MR. J. F., Department of Mathematics, Trinity College, Hartford, Connecticut.
ZOCH, MR. RICHMOND T., Cosmos Club, 1520 H Street, Washington, D. C.

w

s.

a,

go,

ta,

rk,

er-

an,

oit,

ago,

cut.